

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2003, 10:38:11 ; Search time 10 Seconds

(without alignments)
49.772 Million cell updates/sec

Title: US-09-551-151A-43
Perfect score: 64
Sequence: 1 SFQGIAGORNFN 12

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 500 summaries

Database: SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	65.6	779	1	CA11_BOVIN
2	42	65.6	1453	1	CA11_CHICK
3	42	65.6	1453	1	CA11_MOUSE
4	42	65.6	1460	1	CA11_CANFA
5	42	65.6	1464	1	CA11_HUMAN
6	40	62.5	1418	1	CA12_HUMAN
7	40	62.5	1459	1	CA12_MOUSE
8	39	60.9	1320	1	PUR4_NEIMA
9	39	60.9	1320	1	PUR4_NEIMA
10	38.5	60.2	619	1	SELB_HAETN
11	37	57.8	777	1	BISC_ECOLI
12	37	57.8	777	1	CAFM_YERPE
13	36	56.2	258	1	CA14_DROME
14	36	56.2	324	1	UBP2_MOUSE
15	36	56.2	861	1	GUGB_SOLTU
16	36	56.2	919	1	ANDR_HUMAN
17	36	56.2	1334	1	PUR4_DROME
18	36	56.2	1775	1	CA14_DROME
19	35	54.7	353	1	UBP2_MOUSE
20	35	54.7	451	1	ECFE_YERPE
21	35	54.7	605	1	UBP2_MOUSE
22	35	54.7	884	1	ANDR_EULFC
23	35	54.7	895	1	ANDR_MACFA
24	35	54.7	895	1	ANDR_PAPHA
25	35	54.7	907	1	ANDR_CANFA
26	35	54.7	911	1	ANDR_PANTR
27	35	54.7	1882	1	POL2_TRSVR
28	34	53.1	97	1	ECFA_HUMAN
29	34	53.1	266	1	STR_KUEPN
30	34	53.1	278	1	YJ3C_SCHPO
31	34	53.1	283	1	YHMO_YEAST
32	34	53.1	412	1	OBSD_DROME
33	34	53.1	579	1	IVRC_MERTH

34	34	53.1	594	1	DHBK_LYCES
35	34	53.1	595	1	SNX9_HUMAN
36	34	53.1	644	1	LEU2_RHIRA
37	34	53.1	654	1	TPE2_HUMAN
38	34	53.1	689	1	CA29_HUMAN
39	34	53.1	734	1	YD2B_SCHPO
40	34	53.1	773	1	LEU2_USRMA
41	34	53.1	819	1	STL_CHLMO
42	34	53.1	819	1	STL_CHLMO
43	34	53.1	1559	1	STL_CHLMO
44	33	51.6	208	1	YMO6_EMENTI
45	33	51.6	217	1	RS5_MYCLE
46	33	51.6	264	1	HBM4_SYND7
47	33	51.6	303	1	CRKL_HUMAN
48	33	51.6	303	1	CRKL_HUMAN
49	33	51.6	324	1	GSHB_ANASP
50	33	51.6	397	1	AMPQ_PSEAE
51	33	51.6	408	1	APG1_ARCFU
52	33	51.6	522	1	HMC5_CHICK
53	33	51.6	564	1	CDC7_MOUSE
54	33	51.6	574	1	CDC7_MOUSE
55	33	51.6	585	1	CDC7_MOUSE
56	33	51.6	628	1	YK05_CABEL
57	33	51.6	636	1	ABFA_ASPNG
58	33	51.6	642	1	CA13_RAT
59	33	51.6	742	1	TBX3_HUMAN
60	33	51.6	820	1	GLGB_ORYSA
61	33	51.6	931	1	EMRI_MOUSE
62	33	51.6	1049	1	CA13_BOVIN
63	33	51.6	1262	1	CA13_CHICK
64	33	51.6	1464	1	CA13_MOUSE
65	33	50.0	129	1	CA13_HUMAN
66	32	50.0	170	1	Y959_TREPA
67	32	50.0	183	1	MPA5_BOVIN
68	32	50.0	189	1	AR21_CABEL
69	32	50.0	218	1	PGHD_RAT
70	32	50.0	240	1	Y522_HABIN
71	32	50.0	297	1	SMS1_MOUSE
72	32	50.0	385	1	YFCH_ECOLI
73	32	50.0	440	1	ACKA_MCTOC
74	32	50.0	440	1	PCXA_SYNT3
75	32	50.0	442	1	GAG_VILV
76	32	50.0	442	1	GAG_VILV1
77	32	50.0	442	1	GAG_VILV2
78	32	50.0	478	1	2230_HUMAN
79	32	50.0	520	1	FLDA_PSEAE
80	32	50.0	520	1	HMC5_CRIGR
81	32	50.0	520	1	HMC5_HUMAN
82	32	50.0	522	1	HMC5_RAT
83	32	50.0	543	1	SR54_SCHPO
84	32	50.0	551	1	NCAP_SVAL
85	32	50.0	562	1	FXFN_AZCCA
86	32	50.0	658	1	ES22_MOUSE
87	32	50.0	814	1	TPE2_XENLA
88	32	50.0	959	1	IF39_HUMAN
89	32	50.0	1027	1	N100_YEAST
90	32	50.0	1077	1	CAFE_RIFPA
91	32	50.0	1358	1	XYNV_CLOTH
92	32	50.0	1496	1	SIR4_YEAST
93	32	50.0	1527	1	CA25_HUMAN
94	32	50.0	1763	1	CA1H_MOUSE
95	32	50.0	1805	1	CA24_ASCSU
96	32	50.0	2481	1	FINC_XENLA
97	31.5	49.2	827	1	AD17_MOUSE
98	31.5	49.2	837	1	AD17_MOUSE
99	31.5	49.2	2512	1	POLN_STNDV
100	31.5	49.2	2514	1	POLN_STNDV
101	31	48.4	56	1	GP59_BSPPI
102	31	48.4	110	1	CUL5_HUMAN
103	31	48.4	164	1	CUL5_HUMAN
104	31	48.4	190	1	MFAS_MOUSE
105	31	48.4	199	1	ATRC_SYNT3
106	31	48.4	217	1	TDBX_BRUMA
					COBH_SYNT3

004059	lycoperico
0955x1	homo sapien
P17279	rhizomicro
P15923	homo sapien
014055	homo sapien
013718	schizosacch
P49601	uscllago ma
09PK14	chlamydia m
084211	chlamydia t
000681	emericella
050659	mycobacteri
033000	mycobacteri
P42452	synecococc
P46109	homo sapien
P47941	mus musculu
P54580	anubiena sp
P24735	pseudomonas
028523	archaeoglob
P23228	gallus gall
Q82010	mus musculu
000311	homo sapien
P34253	caenorhabd1
P42254	aspergillus
P13943	rattus norv
015119	homo sapien
001401	oryza sativ
001549	mus musculu
P04258	bos taurus
P12121	gallus gall
P08121	mus musculu
P02461	homo sapien
083925	treponema p
Q28022	bos taurus
Q9XVJ3	caenorhabd1
P22057	rattus norv
Q57256	haemophilus
Q9d8y0	mus musculu
P77775	yeich.ecoli
P67275	mycobacteri
P75028	mycobacteri
P03332	visna lent1
P23424	visna lent1
Q23425	visna lent1
091460	homo sapien
033421	pseudomonas
P13704	cricketu
001581	homo sapien
P17425	rattus norv
P21565	schizosacch
P98008	azotobact
064176	mus musculu
001978	xenopus lae
P55881	homo sapien
000269	saccharomyc
P30754	rattus norv
P15154	clostridium
P11978	saccharomyc
P05997	homo sapien
P39001	mus musculu
P27393	ascaris suu
P21263	rattus norv
Q91740	xenopus lae
092108	mus musculu
092108	rattus norv
P03317	sinobis vit
P27283	sinobis vit
048413	bacterioph
P58622	homo sapien
099216	mus musculu
P73869	synecococc
017172	brucella mala
P74304	synecococc

107	31	48.4	248	1	GSPN_ERWCA	P31710	erwinia car	180	30	46.9	279	1	ATPG_MYCPN	Q50330	mycoplasma
108	31	48.4	264	1	FLGC_BACSU	P23446	bacillus su	181	30	46.9	308	1	Y04P_MYCTU	Q10786	mycobacteri
109	31	48.4	265	1	RREP_VSVSU	P03520	vesicular s	182	30	46.9	309	1	KHSE_THERVO	Q979x5	thermoplasm
110	31	48.4	269	1	Y0E1_ECOLI	Q46542	eschcheria	183	30	46.9	343	1	YOKA_BACSU	P54564	bacillus su
111	31	48.4	298	1	OCCR_AGRTR	Q00679	agrobacteri	184	30	46.9	345	1	KOKA_EUGGR	P06651	euglena gra
112	31	48.4	304	1	YD24_MYCTU	Q10636	mycobacteri	185	30	46.9	352	1	AMBP_HUMAN	P02760	homo sapien
113	31	48.4	312	1	VG06_BPM5	Q05278	mycobacteri	186	30	46.9	352	1	PSBA_CHLEL	P35860	chlorella e
114	31	48.4	319	1	MYOD_SHEEP	P29331	ovis aries	187	30	46.9	352	1	PSBA_CHLEL	P07753	chlamydomon
115	31	48.4	321	1	VG06_BPM2	Q46402	mycobacteri	188	30	46.9	352	1	PSBA_PEA	P06585	pisum sativ
116	31	48.4	339	1	MDR_METFE	P61422	methanother	189	30	46.9	352	1	PSBA_PETHY	P04999	petunia hyb
117	31	48.4	356	1	HISB_ECOLI	P06586	eschcheria	190	30	46.9	352	1	PSBA_PICAB	PS0155	fices abies
118	31	48.4	366	1	CAS4_BPHAM	P18503	ephydatia m	191	30	46.9	352	1	PSBA_PINTH	P24664	finus thunb
119	31	48.4	410	1	APGM_PPRAB	Q972m6	pyrococcus	192	30	46.9	352	1	PSBA_PORDE	P36461	populus del
120	31	48.4	411	1	DHE3_MAIZE	Q43260	zea mays (m	193	30	46.9	352	1	PSBA_SECE	P10510	secale cere
121	31	48.4	412	1	ARGM_PPRHO	O57742	pyrococcus	194	30	46.9	352	1	PSBA_SINAL	P11848	sinapis alb
122	31	48.4	413	1	E2P4_HUMAN	Q16254	homo sapien	195	30	46.9	352	1	PSBA_SIMAL	P02957	glycine max
123	31	48.4	482	1	Y269_TREPA	Q83293	treponema p	196	30	46.9	352	1	PSBA_SIOG	P27201	spirodela o
124	31	48.4	483	1	EXON_HSVSA	Q01013	herpesvirus	197	30	46.9	352	1	PSBA_SPIOL	PSBA_PAMAY	
125	31	48.4	493	1	Y911_CAUCR	P33976	caulobacter	198	30	46.9	353	1	PSBA_ARATH	PSBA_AMAHY	
126	31	48.4	524	1	SPAI_STRAU	P02976	staphylococ	199	30	46.9	353	1	PSBA_BRANA	PSBA_ARYTH	
127	31	48.4	561	1	EST4_RAT	Q64573	rattus norv	200	30	46.9	353	1	PSBA_CHIMO	PSBA_BRANA	
128	31	48.4	561	1	EST5_RAT	Q63010	rattus norv	201	30	46.9	353	1	PSBA_CHIMU	PSBA_CHIMO	
129	31	48.4	575	1	YUWA_BACSU	P43861	bacillus su	202	30	46.9	353	1	PSBA_CONCI	PSBA_CHIMU	
130	31	48.4	591	1	IF37_ARATH	P56820	arabidopsis	203	30	46.9	353	1	PSBA_CONSU	PSBA_CONCI	
131	31	48.4	603	1	BGLR_ECOLI	P05804	eschcheria	204	30	46.9	353	1	PSBA_CONSU	PSBA_CONSU	
132	31	48.4	626	1	TESK_HUMAN	Q15569	homo sapien	205	30	46.9	353	1	PSBA_CUSRE	PSBA_CUSRE	
133	31	48.4	709	1	ANDR_RABIT	P49699	oryctolagus	206	30	46.9	353	1	PSBA_DUMHI	PSBA_DUMHI	
134	31	48.4	718	1	FLGE_HELPJ	Q924Y0	hellicobacte	207	30	46.9	353	1	PSBA_HORVU	PSBA_HORVU	
135	31	48.4	718	1	FLGE_HELPJ	P50610	hellicobacte	208	30	46.9	353	1	PSBA_HORVU	PSBA_HORVU	
136	31	48.4	764	1	AKRI_YEAST	P39010	saccharomyc	209	30	46.9	353	1	PSBA_MARZO	PSBA_MARZO	
137	31	48.4	869	1	SUS2_PEA	O24301	pisum sativ	210	30	46.9	353	1	PSBA_MESY	PSBA_MESY	
138	31	48.4	869	1	CFAC_ECOLI	P25733	eschcheria	211	30	46.9	353	1	PSBA_MESY	PSBA_MESY	
139	31	48.4	899	1	EMBA_MYCSM	O50394	mycobacteri	212	30	46.9	353	1	PSBA_MESY	PSBA_MESY	
140	31	48.4	1092	1	RROD_ROTGI	P35542	rotavirus (213	30	46.9	353	1	PSBA_MESY	PSBA_MESY	
141	31	48.4	1286	1	ALDO_ECOLI	Q03155	eschcheria	214	30	46.9	353	1	PSBA_PROHO	PSBA_PROHO	
142	31	48.4	1319	1	MNI_HUMAN	Q10571	homo sapien	215	30	46.9	353	1	PSBA_TOBAC	PSBA_TOBAC	
143	31	48.4	1391	1	PRAX_MOUSE	Q05103	mus musculu	216	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
144	31	48.4	1618	1	NEST_HUMAN	P48681	homo sapien	217	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
145	31	48.4	1669	1	CA14_HUMAN	P02462	homo sapien	218	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
146	31	48.4	1669	1	CA14_MOUSE	P02463	mus musculu	219	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
147	31	48.4	1744	1	RPA1_HUMAN	P16355	trypanosoma	220	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
148	31	48.4	1793	1	LMB2_HYRBB	P35268	homo sapien	221	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
149	31	48.4	1892	1	YB35_RICCN	Q924nd6	ricekeltia	222	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
150	31	48.4	2095	1	RRLP_TOSV	P37800	toscana vir	223	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
151	31	48.4	2524	1	NORC_XENLA	P21783	xenopus lae	224	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
152	31	48.4	2569	1	LM43_MOUSE	Q61783	mus musculu	225	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
153	31	48.4	3034	1	CIH1_MOUSE	Q31561	mus musculu	226	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
154	31	48.4	3034	1	ACR1_LONAC	P23804	lonomia ach	227	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
155	30.5	47.7	379	1	DNAL_STRAU	P43555	staphylococ	228	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
156	30.5	47.7	518	1	FETA_HSV11	P04290	herpes slmp	229	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
157	30.5	47.7	609	1	FEET_HORSE	P49066	equus cabal	230	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
158	30.5	47.7	775	1	PLSR_MYCLE	Q9X7B0	mycobacteri	231	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
159	30.5	47.7	1103	1	VG37_BPARI	Q908D5	bacterioph	232	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
160	30.5	47.7	1509	1	MYSN_ACACA	P05559	acanthamob	233	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
161	30	46.9	87	1	ATPK_BOVIN	Q28851	bos taurus	234	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
162	30	46.9	87	1	ATPK_PIG	Q95339	sus scrofa	235	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
163	30	46.9	93	1	IM10_EMENT	Q9Y8A8	emeritella	236	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
164	30	46.9	102	1	CPAI_CANPG	P81575	cancer pagu	237	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
165	30	46.9	123	1	GLN2_METBA	P54809	methanosarc	238	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
166	30	46.9	127	1	SPH1_BACHD	Q9K859	metanobact	239	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
167	30	46.9	139	1	LPOY_MYCTU	O53412	mycobacteri	240	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
168	30	46.9	156	1	NRDL_STRPN	Q97103	strepococ	241	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
169	30	46.9	164	1	PHEA_FREDI	P05098	freymyella d	242	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
170	30	46.9	164	1	PHEA_PSEAY	P29296	pseudonaba	243	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
171	30	46.9	164	1	PHEA_SYNY1	P20778	synecocyst	244	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
172	30	46.9	167	1	R55_BUCAI	P57574	buchnera ap	245	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
173	30	46.9	181	1	RK5_PORPU	P51302	porphyra pu	246	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
174	30	46.9	181	1	SYDE_ECOLI	P43526	eschcheria	247	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
175	30	46.9	202	1	Y038_MYCLE	Q50191	mycobacteri	248	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
176	30	46.9	246	1	YLS4_CORGL	P94338	corynebacte	249	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
177	30	46.9	251	1	PMGY_TREPA	P96121	treponema p	250	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
178	30	46.9	255	1	CBPM_STRAU	P00731	streptomyc	251	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	
179	30	46.9	279	1	ATPG_MYCCE	P47640	mycoplasma	252	30	46.9	353	1	PSBA_VICIN	PSBA_VICIN	

253	30	46.9	430	1	SHU2_ECOLI	P09746	eschertichia	326	30	46.9	3079	1	IRA2_YEAST	P19158	saccharomyc
254	30	46.9	433	1	SHU6_ECOLI	P09747	eschertichia	327	30	46.9	3313	1	CUR3_RAT	O88278	rattus norv
255	30	46.9	435	1	SHU2_SCHPO	P40234	schizosacch	328	30	46.9	3587	1	SRR1_BACU	P27206	bacillus su
256	30	46.9	442	1	SHU3_ECOLI	P09747	eschertichia	329	30	46.9	3587	1	SRR2_BACU	O04747	bacillus su
257	30	46.9	444	1	SHU4_ECOLI	P09748	eschertichia	330	30	46.9	4451	1	GRSB_BACR	P14688	b gramicid
258	30	46.9	444	1	SHU5_ECOLI	P09751	eschertichia	331	30	46.1	685	1	SYG_TREPA	O83678	treponema p
259	30	46.9	456	1	SHU7_ECOLI	O10614	mycobacteri	332	29.5	46.1	1505	1	ERR2_YEAS	P05455	saccharomyc
260	30	46.9	456	1	YC88_MYCTU	O91yuh	mus musculu	333	29.5	46.1	685	1	CUT2_HUMAN	O14520	homo sapien
261	30	46.9	470	1	SSP1_MOUSE	O9n85	homo sapien	334	29	45.3	103	1	CHLB_NEPX	P17850	homo sapien
262	30	46.9	473	1	SSP1_MOUSE	O9n85	homo sapien	335	29	45.3	114	1	IKO5_HCWA	P16803	human cytom
263	30	46.9	474	1	SHU1_ECOLI	O46078	corynebacte	336	29	45.3	124	1	SPEH_THERN	O8474	thermoaer
264	30	46.9	475	1	KPKY_CORGL	O75688	homo sapien	337	29	45.3	128	1	SPEH_BACU	O34426	bacillus su
265	30	46.9	479	1	P2CB_HUMAN	O60989	mus musculu	338	29	45.3	140	1	LYSA_DROME	P37157	bacillus su
266	30	46.9	496	1	BIR4_MOUSE	O9-016	rattus norv	339	29	45.3	140	1	LYSB_DROME	O08694	drosophila
267	30	46.9	496	1	BIR4_MOUSE	P12206	gallus gall	340	29	45.3	140	1	LYSE_DROME	P37159	drosophila
268	30	46.9	503	1	CA19_CHICK	P03819	eschertichia	341	29	45.3	183	1	RS5_MICU	P33105	micrococcus
269	30	46.9	510	1	CEAB_ECOLI	P22632	zymomonas m	342	29	45.3	185	1	YCT1_BACFI	O04453	bacillus fl
270	30	46.9	511	1	INVA_ZYMO	P35636	zymomonas m	343	29	45.3	214	1	ATT_BOMMO	O26431	bombyx mori
271	30	46.9	512	1	NAB2_YEAST	P32505	saccharomyc	344	29	45.3	214	1	TESC_HUMAN	O96b82	homo sapien
272	30	46.9	525	1	TF65_MOUSE	O04207	mus musculu	345	29	45.3	217	1	SM5H_DROME	O91j36	drosophila
273	30	46.9	554	1	FM04_RABIT	O05254	mycobacteri	346	29	45.3	221	1	THIE_PASNU	P37930	pasteurella
274	30	46.9	554	1	FM04_RABIT	P28249	eschertichia	347	29	45.3	228	1	YF77_LISMO	O81674	listeria mo
275	30	46.9	595	1	DPOL_BPM15	P40534	saccharomyc	348	29	45.3	239	1	YG12_LISTN	O92b08	homo sapien
276	30	46.9	617	1	ASMA_ECOLI	P02454	rattus norv	349	29	45.3	240	1	SMS2_MOUSE	O9d471	mus musculu
277	30	46.9	656	1	YID7_YEAST	O01819	drosophila	350	29	45.3	247	1	YOFO_BACU	P54474	bacillus su
278	30	46.9	671	1	CA11_RAT	O08838	rattus norv	351	29	45.3	252	1	AGL6_ARATH	P29336	arabidopsis
279	30	46.9	682	1	CONN_DROME	P79101	bos taurus	352	29	45.3	259	1	DHSB_PANDE	O39662	arabidopsis
280	30	46.9	684	1	AMPH_RAT	O9nKf6	homo sapien	353	29	45.3	269	1	DAPH_VIBCH	O9nKf7	vibrio chol
281	30	46.9	684	1	CPEC_BOVIN	O9nKf6	homo sapien	354	29	45.3	270	1	LPSC_RHIME	O9nKf9	rhizobium m
282	30	46.9	684	1	CPEC_HUMAN	P17998	eschertichia	355	29	45.3	279	1	PHEA_LACIA	O9c612	lactococcus
283	30	46.9	684	1	CPEC_MOUSE	P54480	bacillus su	356	29	45.3	285	1	ATPG_BACME	P02602	bacillus me
284	30	46.9	697	1	YQGF_BACU	P18250	phycomycs	357	29	45.3	287	1	ATPG_BACR3	P09222	bacillus ps
285	30	46.9	716	1	SYGB_SYNX3	O9j3lt	homo sapien	358	29	45.3	287	1	ATPG_BACCA	P412007	bacillus ca
286	30	46.9	722	1	LEU2_PHYBL	P55811	rhizopus ni	359	29	45.3	291	1	ATPG_BACST	P21007	bacillus st
287	30	46.9	744	1	LEU2_PHYBL	P55251	rhizomucor	360	29	45.3	294	1	ARGC_ECOLI	O07906	bacillus st
288	30	46.9	750	1	LEU2_RHINI	P13876	influenza b	361	29	45.3	301	1	ADT_ANOGA	P23228	asporogallus
289	30	46.9	750	1	LEU2_RHINI	O36431	influenza b	362	29	45.3	307	1	CC36_CAREL	P35887	lactobacilli
290	30	46.9	755	1	LEU2_RHINI	P39016	saccharomyc	363	29	45.3	314	1	REP_LACHT	P35887	lactobacilli
291	30	46.9	770	1	RRP3_INBAC	P38970	saccharomyc	364	29	45.3	320	1	GSHB_STYNT	P37433	synechococc
292	30	46.9	770	1	RRP3_INBAC	O9y2k6	homo sapien	365	29	45.3	323	1	GSHB_STYNT	O32433	synechococc
293	30	46.9	770	1	RRP3_INBAC	P20848	homo sapien	366	29	45.3	339	1	ARGC_LACIA	O06052	mycobacteri
294	30	46.9	770	1	RRP3_INBAC	P52948	homo sapien	367	29	45.3	340	1	ARGC_LACIA	O06052	mycobacteri
295	30	46.9	834	1	MRP5_YEAST	P43793	rattus norv	368	29	45.3	341	1	PLSX_ANASP	O08318	lactobacilli
296	30	46.9	855	1	HALE_YEAST	O08162	saccharomyc	369	29	45.3	344	1	POPA_RALSO	O82053	arabidopsis
297	30	46.9	913	1	UBPK_HUMAN	O04857	mus musculu	370	29	45.3	345	1	PRSG_BACSU	O9rdb0	talisconia s
298	30	46.9	921	1	CA19_HUMAN	O06330	saccharomyc	371	29	45.3	346	1	ARGC_BACU	P08300	bacillus su
299	30	46.9	937	1	NU98_HUMAN	O98632	rice dwarf	372	29	45.3	347	1	TDH_BACSU	P23715	bacillus su
300	30	46.9	937	1	NU98_HUMAN	O98632	rice dwarf	373	29	45.3	347	1	TDH_BACSU	O31776	bacillus su
301	30	46.9	1001	1	RR4_YEAST	O52236	treponema p	374	29	45.3	350	1	GANA_ASPAC	P48862	asporogallus
302	30	46.9	1025	1	CA16_MOUSE	P81122	mus musculu	375	29	45.3	359	1	ODH_HERSE	P77875	herbaspirill
303	30	46.9	1113	1	VP2_YEAST	O9y4n2	homo sapien	376	29	45.3	368	1	PHA2_YEAS	O44237	arctobacte
304	30	46.9	1116	1	VP2_YEAST	O9y4n2	homo sapien	377	29	45.3	368	1	PHA2_YEAS	P32452	saccharomyc
305	30	46.9	1127	1	Y855_TREPA	O4350	tana catesb	378	29	45.3	369	1	TRMA_VIBCH	O9wtv6	mus musculu
306	30	46.9	1201	1	MRD_MYXXA	O93484	oncofichu	379	29	45.3	381	1	DICH_DROME	O9kvi0	vibrio chol
307	30	46.9	1321	1	IR82_MOUSE	P70298	mus musculu	380	29	45.3	382	1	DICH_DROME	O24533	drosophila
308	30	46.9	1324	1	IR82_MOUSE	P28666	mus musculu	381	29	45.3	383	1	CYCR_CHRVI	O82967	chromatium
309	30	46.9	1355	1	IR82_MOUSE	P34881	arabidopsis	382	29	45.3	383	1	DGT1_RICPR	O9ze62	trickettsia
310	30	46.9	1426	1	CA21_ONCWY	O92072	gallus gall	383	29	45.3	384	1	WR25_ARATH	O92j13	trickettsia
311	30	46.9	1451	1	AZM2_MOUSE	P10071	homo sapien	384	29	45.3	406	1	CG16_HUMAN	O22921	arabidopsis
312	30	46.9	1506	1	PK3G_MOUSE	O27746	paracentrot	385	29	45.3	411	1	HMHD_PYRAB	O9v1r3	pyrococcus
313	30	46.9	1534	1	DNN1_MOUSE	P13864	mus musculu	386	29	45.3	413	1	GAG2_HUMAN	P10264	homo sapien
314	30	46.9	1537	1	DNN1_MOUSE	O92330	rattus norv	387	29	45.3	415	1	HMHD_PYRHO	O59469	pyrococcus
315	30	46.9	1596	1	DNN1_MOUSE	O61245	mus musculu	388	29	45.3	418	1	SSXT_MOUSE	P19223	rattus norv
316	30	46.9	1612	1	DNN1_MOUSE	P12107	homo sapien	389	29	45.3	426	1	PYRC_THENO	O62280	mus musculu
317	30	46.9	1616	1	DNN1_MOUSE	P51805	homo sapien	390	29	45.3	426	1	TWIN_DROME	P96081	thermus aqu
318	30	46.9	1620	1	DNN1_MOUSE	O53553	mycobacteri	391	29	45.3	434	1	CBPS_STRCP	O03019	drosophila
319	30	46.9	1620	1	DNN1_MOUSE	P98163	drosophila	392	29	45.3	463	1	SAHH_CAUCR	P39041	strepomyce
320	30	46.9	1804	1	CA1B_MOUSE	P33453	ukunleml v	393	29	45.3	466	1	DCEA_ECO57	O9abn0	caulobacter
321	30	46.9	1806	1	CA1B_MOUSE			394	29	45.3	466	1	DCEA_ECO57	P58228	eschertichia
322	30	46.9	1871	1	CA1B_MOUSE			395	29	45.3	466	1	DCEA_ECO57	P80063	eschertichia
323	30	46.9	1901	1	YL_DROME			396	29	45.3	466	1	DCEA_ECO57		
324	30	46.9	1984	1	YL_DROME			397	29	45.3	466	1	DCEA_ECO57		
325	30	46.9	2103	1	RRPL_UUK			398	29	45.3	466	1	DCEA_ECO57		

399	29	45.3	466	1	DCBB_ECOLI	P28302	eschlerichia	472	29	45.3	1858	1	P3K2_DICDI	P54674	dictyosteli
400	29	45.3	471	1	COLO_TORMA	003637	torpedo mar	473	29	45.3	1926	1	LPH_RABIT	P09849	oryctolagus
401	29	45.3	472	1	COBO_PYRKO	034475	pyrococcus	474	29	45.3	2164	1	POLG_HRV9	P07210	human rhino
402	29	45.3	481	1	BIND_STRPU	P06651	strongyloce	475	29	45.3	2175	1	HMCU_DROME	P10180	drosophila
403	29	45.3	483	1	MRCO_MESAU	09wud9	mesocricetu	476	29	45.3	2386	1	FINC_DROME	P02751	homo sapien
404	29	45.3	486	1	RBL1_RHOSH	P27997	rhodobacter	477	29	45.3	2471	1	NTC2_HUMAN	P04721	homo sapien
405	29	45.3	488	1	GATA_THETN	08rc40	thermoanaer	478	29	45.3	2471	1	NTC2_HUMAN	P04721	homo sapien
406	29	45.3	490	1	YAM7_SCHPO	010062	schizosacch	479	29	45.3	2549	1	FRAP_MOUSE	P09430	rattus norv
407	29	45.3	493	1	CGAD_DROME	09v4u9	drosophila	480	29	45.3	2944	1	CA17_HUMAN	P02388	homo sapien
408	29	45.3	495	1	PURI_SYNY3	055621	synchocyst	481	29	45.3	3014	1	CLRI_HUMAN	P02388	homo sapien
409	29	45.3	507	1	CRNA_EMENTI	P22152	emericeella	482	29	45.3	3137	1	CA36_CHICK	P19889	gallus gall
410	29	45.3	515	1	COAT_TRSV	088894	tobacco rin	483	29	45.3	3172	1	ERY3_SACR	P03133	saccharopol
411	29	45.3	518	1	FUS_MOUSE	P56929	mus musculu	484	29	45.3	3176	1	CA36_HUMAN	P12111	homo sapien
412	29	45.3	519	1	AT13_ARATH	09s409	arabidopsis	485	29	45.3	3491	1	ERY1_SACR	P03131	saccharopol
413	29	45.3	523	1	CP5E_CANMA	P31433	eschlerichia	486	29	45.3	3511	1	MY15_MOUSE	P09424	mus musculu
414	29	45.3	523	1	CP5E_CANMA	P24458	candida mal	487	29	45.3	4967	1	RYR2_HUMAN	P02778	homo sapien
415	29	45.3	526	1	CA21_RABIT	028668	oryctolagus	488	29	45.3	4969	1	RYR2_RABIT	P30957	oryctolagus
416	29	45.3	526	1	FUS_HUMAN	P35637	homo sapien	489	29	45.3	4969	1	RYR2_HUMAN	P30957	oryctolagus
417	29	45.3	534	1	CHIT_MANSE	P35362	manduca sex	490	29	45.3	4969	1	RYR2_HUMAN	P30957	oryctolagus
418	29	45.3	556	1	TCPE_TETPY	P54409	tetrahymena	491	29	45.3	4969	1	RYR2_HUMAN	P30957	oryctolagus
419	29	45.3	556	1	MBHM_ECOLI	P37181	eschlerichia	492	29	45.3	4969	1	RYR2_HUMAN	P30957	oryctolagus
420	29	45.3	567	1	EST1_HUMAN	P23141	homo sapien	493	29	45.3	4969	1	RYR2_HUMAN	P30957	oryctolagus
421	29	45.3	569	1	YICH_ECOLI	P31433	eschlerichia	494	29	45.3	4969	1	RYR2_HUMAN	P30957	oryctolagus
422	29	45.3	588	1	CYCD_ECOLI	P29018	eschlerichia	495	29	45.3	4969	1	RYR2_HUMAN	P30957	oryctolagus
423	29	45.3	607	1	KITH_EBV	P03177	epstein-bar	496	29	45.3	4969	1	RYR2_HUMAN	P30957	oryctolagus
424	29	45.3	623	1	CA44_RABIT	P57187	oryctolagus	497	29	45.3	4969	1	RYR2_HUMAN	P30957	oryctolagus
425	29	45.3	649	1	TFE2_MESAU	P98180	mesocricetu	498	29	45.3	4969	1	RYR2_HUMAN	P30957	oryctolagus
426	29	45.3	649	1	TFE2_RAT	P21677	rattus norv	499	29	45.3	4969	1	RYR2_HUMAN	P30957	oryctolagus
427	29	45.3	653	1	SCAG_RABIT	028738	oryctolagus	500	29	45.3	4969	1	RYR2_HUMAN	P30957	oryctolagus
428	29	45.3	687	1	WHIT_DROME	P10090	drosophila								
429	29	45.3	693	1	LYS4_YEAST	P49367	saccharomyc								
430	29	45.3	730	1	SMIA_SCHAM	026473	schistocerc								
431	29	45.3	741	1	MAS2_MYCTU	050596	mycobacteri								
432	29	45.3	741	1	TRM3_HUMAN	075382	homo sapien								
433	29	45.3	744	1	TRM3_MOUSE	091172	mus musculu								
434	29	45.3	744	1	TRM3_RAT	070277	rattus norv								
435	29	45.3	746	1	ABP_RAT	P36633	rattus norv								
436	29	45.3	749	1	PEX_HUMAN	P78662	homo sapien								
437	29	45.3	749	1	PEX_MOUSE	P70669	mus musculu								
438	29	45.3	750	1	ELS_CHICK	P07916	gallus gall								
439	29	45.3	751	1	ABP_HUMAN	P19801	homo sapien								
440	29	45.3	758	1	PKAI_YEAST	P41909	saccharomyc								
441	29	45.3	774	1	GLVB_PLAFG	P02895	plasmodium								
442	29	45.3	819	1	FVB_MOUSE	035601	mus musculu								
443	29	45.3	838	1	P_HUMAN	004671	homo sapien								
444	29	45.3	861	1	GAL4_YEAST	P04386	saccharomyc								
445	29	45.3	902	1	ANDR_RAT	P15207	rattus norv								
446	29	45.3	911	1	CA1B_BOVIN	028083	bos taurus								
447	29	45.3	918	1	YMUJ_CAEEL	P34487	caenorhabdi								
448	29	45.3	926	1	POBI_YEAST	001454	saccharomyc								
449	29	45.3	997	1	YPR2_CAEEL	020256	caenorhabdi								
450	29	45.3	1093	1	YKDS_CAEEL	003363	caenorhabdi								
451	29	45.3	1108	1	UBPB_SCHPO	009879	schizosacch								
452	29	45.3	1143	1	CA1I_HUMAN	014993	homo sapien								
453	29	45.3	1176	1	HMDH_PHYBL	012649	phycomyc								
454	29	45.3	1210	1	EGFR_MOUSE	001279	mus musculu								
455	29	45.3	1333	1	ADO_RAT	0920u5	rattus norv								
456	29	45.3	1334	1	ADO_RABIT	P80456	oryctolagus								
457	29	45.3	1364	1	CA21_BOVIN	P02465	bos taurus								
458	29	45.3	1366	1	CA21_CANFA	046392	canis famill								
459	29	45.3	1366	1	CA21_HUMAN	P08123	homo sapien								
460	29	45.3	1372	1	CA21_MOUSE	001149	mus musculu								
461	29	45.3	1372	1	CA21_RAT	P02466	rattus norv								
462	29	45.3	1480	1	PANI_YEAST	P32521	saccharomyc								
463	29	45.3	1567	1	RW1_DROME	09v7h4	drosophila								
464	29	45.3	1603	1	CA1F_HUMAN	007092	homo sapien								
465	29	45.3	1650	1	CA2B_MOUSE	064739	mus musculu								
466	29	45.3	1712	1	CA24_HUMAN	P08572	homo sapien								
467	29	45.3	1736	1	CA2B_HUMAN	P17140	caenorhabdi								
468	29	45.3	1758	1	CA24_CAEEL	061292	mus musculu								
469	29	45.3	1799	1	LMB2_MOUSE	P15600	rattus norv								
470	29	45.3	1801	1	FAS2_RAT	P15368	p fatty aci								
471	29	45.3	1857	1	FAS2_PENPA										

RESULT 1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																												
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

ALIGNMENTS

RESULT 1

ID	CALL_BOVIN	STANDARD	PRT	779 AA.
AC	P02453			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Collagen alpha 1(I) chain (Fragments).			
GN	COL1A1.			
OS	Bos taurus (Bovine).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;			
OC	Bovidae; Bovinae; Bos.			
OX	NCBI_TaxID=9913;			
RP	SEQUENCE OF 1-19.			
RA	MEDLINE=72255334; PubMed=4115172;			
RA	Rautenberg J., Timpl R., Furtmayr H.;			
RT	"Structural characterization of N-terminal antigenic determinants in			
RT	calf and human collagen.";			
RL	Eur. J. Biochem. 27:231-237(1972).			
RL	[2]			
RP	SEQUENCE OF 20-145.			
RX	MEDLINE=76022320; PubMed=1164916;			
RA	Fietzek P.P., Kuehn K.;			
RT	"The covalent structure of collagen: amino-acid sequence of the			
RT	cyanogen-dromide peptides alpha-1-CB2, alpha-1-CB4 and alpha-1-CB5			
RT	from calf-skin collagen.";			
RL	Eur. J. Biochem. 52:77-82(1975).			
RL	[3]			
RP	SEQUENCE OF 146-294.			
RX	MEDLINE=73049499; PubMed=4673951;			
RA	Fietzek P.P., Wendt P., Kell I., Kuehn K.;			
RT	"The covalent structure of collagen: amino acid sequence of alpha-1-			
RT	CB3 from calf skin collagen.";			
RL	FEBS Lett. 26:74-76(1972).			
RL	[4]			
RP	SEQUENCE OF 295-562.			
RX	MEDLINE=74086118; PubMed=4359390;			

P54674 dictyosteli
P09849 oryctolagus
P07210 human rhino
P10180 drosophila
P02751 homo sapien
P04721 homo sapien
P09430 rattus norv
P02388 homo sapien
P19889 gallus gall
P03133 saccharopol
P12110 homo sapien
P03133 saccharopol
P09424 mus musculu
P02778 homo sapien
P30957 oryctolagus
P21810 homo sapien
P02678 canis famill
P28653 mus musculu
P47853 rattus norv
O46390 ovis aries
O46403 equus cabal
P28050 gorilla gor
P02778 pan troglod
P79334 bos taurus
P11217 homo sapien

RA Fietzek P.P., Rexrodt F.W., Hopper K.E., Kuehn K.;
 RT "The covalent structure of collagen. 2. The amino-acid sequence of
 alpha-1-B87 from calf-skin collagen.";
 RL Eur. J. Biochem. 38:396-400(1973).
 RN [5]
 RP SEQUENCE OF 563-675.
 RX MEDLINE=73042276; PubMed=4343808;
 RA Wendt P., Mark K.V.D., Rexrodt F., Kuehn K.;
 RT "The covalent structure of collagen. The amino-acid sequence of the
 112-residues. Amino-terminal part of peptide alpha-1-C86 from calf-
 skin collagen.";
 RL Eur. J. Biochem. 30:169-183(1972).
 RN [6]
 RP SEQUENCE OF 676-751.
 RX MEDLINE=73042275; PubMed=4343807;
 RA Fietzek P.P., Rexrodt F.W., Wendt P., Stark M., Kuehn K.;
 RT "The covalent structure of collagen. Amino-acid sequence of peptide
 alpha-1-C86-C2.";
 RL Eur. J. Biochem. 30:163-168(1972).
 CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 (FIBRILLAR FORMING COLLAGEN).
 CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
 CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
 BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
 HYDROXYAPATITE.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC O-LINKED GLYCAN CONSISTS OF GLC-GAL DISACCHARIDE.
 CC -1- MISCELLANEOUS: THE COMPLETE CHAIN CONTAINS 1052 RESIDUES.
 DR PIR: A91193; C98015.
 DR Interpro: IPR000087; Collagen.
 DR Interpro: IPR001007; VWF_C.
 DR ProDom: PD000007; Collagen; 1.
 DR PROSITE: PS01208; VWF; PARTIAL.
 KM Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen.
 FT MOD_RES 1 1
 FT MOD_RES 9 9
 FT MOD_RES 103 103
 FT CARBOHYD 103 103
 FT MOD_RES 115 115
 FT MOD_RES 124 124
 FT MOD_CONS 145 146
 FT MOD_RES 274 274
 FT MOD_RES 346 346
 FT MOD_RES 424 424
 FT MOD_RES 496 496
 FT MOD_RES 658 658
 FT MOD_RES 670 670
 FT MOD_RES 726 726
 SQ SEQUENCE 779 AA; 70346 MW; E554A7FE084283D1 CRC64;
 Query Match 65.68; Score 42; DB 1; Length 779;
 Best Local Similarity 100.0%; Pred. No. 3.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 POGIAGOR 9
 Db 513 POGIAGOR 520
 RESULT 2
 ID CALL_CHICK STANDARD; PRT; 1453 AA.
 AC P02457;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN COL1A1.

OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE OF 1-153 FROM N.A.
 RX MEDLINE=88056316; PubMed=3678834;
 RA Finer M.H., Boedtker H., Doty P.;
 RT "Construction and characterization of cDNA clones encoding the 5' end
 of the chicken pro alpha 1(I) collagen mRNA.";
 RL Gene 56:71-78(1987).
 RN [2]
 RP SEQUENCE OF 1-144 FROM N.A.
 RX MEDLINE=88007542; PubMed=2820966;
 RA Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
 RT "Unusual DNA sequences located within the promoter region and the
 first intron of the chicken pro-alpha 1(I) collagen gene.";
 RL J. Biol. Chem. 262:13323-13332(1987).
 RN [3]
 RP SEQUENCE OF 152-1187.
 RX MEDLINE=82231995; PubMed=7093229;
 RA Higberger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M.,
 Kang A.H., Gross J.;
 RT "Amino acid sequence of chick skin collagen alpha 1(I)-C88 and the
 complete primary structure of the helical portion of the chick skin
 collagen alpha 1(I) chain.";
 RL Biochemistry 21:2048-2055(1982).
 RN [4]
 RP SEQUENCE OF 1200-1205.
 RX MEDLINE=72243016; PubMed=5047697;
 RA Eyre D.R., Glimcher M.J.;
 RT "Evidence for a previously undetected sequence at the carboxyterminus
 of the alpha 1 chain of chicken bone collagen.";
 RL Biochem. Biophys. Res. Commun. 48:720-726(1972).
 RN [5]
 RP SEQUENCE OF 981-1453 FROM N.A.
 RX MEDLINE=81160715; PubMed=6927845;
 RA Fuller F., Boedtker H.;
 RT "Sequence determination and analysis of the 3' region of chicken pro-
 alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
 including the carboxy-terminal propeptide sequences.";
 RL Biochemistry 20:996-1006(1981).
 RN [6]
 RP SEQUENCE OF 1311-1453 FROM N.A.
 RX MEDLINE=80134546; PubMed=6987088;
 RA Showalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T.,
 Pastan I., Decrombrughe B., Fietzek P.P., Olsen B.R.;
 RT "Nucleotide sequence of a collagen cDNA-fragment coding for the
 carboxyl end of pro alpha 1(I)-chains.";
 RL FEBS Lett. 111:61-65(1980).
 CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 (FIBRILLAR FORMING COLLAGEN).
 CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
 CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
 BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
 HYDROXYAPATITE.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- SIMILARITY: CONTAINS 1 VWF DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC EMBL: M17839; AAA48704.1;
 CC EMBL: M17838; AAA48704.1; JOINED.
 CC EMBL: V00401; CAA23695.1;
 CC EMBL: M10571; AAA48671.1; ALT_SEQ.

EMBL: M17607; AAA48672.1; -
 DR PIR: A02857; CGCHIS.
 DR PIR: A27179; A27179.
 DR PIR: A29367; A29367.
 DR InterPro: IPR000087; Collagen.
 DR InterPro: IPR000885; Fib_collagen_C.
 DR InterPro: IPR001007; VWF_C.
 DR Pfam: PF00093; VWC; 1.
 DR Pfam: PF01391; Collagen; 18.
 DR Pfam: PF01410; COLFI; 1.
 DR Pfam: PF000007; Collagen; 2.
 DR ProDom: PD002078; Fib_collagen_C; 1.
 DR SMART: SM00038; COLFI; 1.
 DR SMART: SM00214; VWC; 1.
 DR PROSITE: PS01208; VWF; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 22
 FT PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
 FT CHAIN 152 1205 COLLAGEN ALPHA 1(I) CHAIN.
 FT PROPEP 1206 1453 C-TERMINAL PROPEPTIDE.
 FT DOMAIN 31 89 VWC.
 FT MOD_RES 152 152 PYRROLIDONE CARBOXYLIC ACID.
 FT MOD_RES 254 254 HYDROXYLATION (POTENTIAL).
 FT MOD_RES 851 851 HYDROXYLATION (POTENTIAL).
 FT MOD_RES 1081 1081 HYDROXYLATION (POTENTIAL).
 FT MOD_RES 1097 1097 HYDROXYLATION (POTENTIAL).
 FT MOD_RES 1153 1153 HYDROXYLATION (ONLY 3-HYDROXYPRO AND THE ONLY HYDROXYLATED PRO IN POSITION X (IN THE G-X-Y UNIT IN THE ALPHA 1(I) CHAIN)).
 FT CONFLICT 1187 1187 F -> L (IN REF. 5).
 FT CONFLICT 1441 1441 Q -> H (IN REF. 6).
 SQ SEQUENCE 1453 AA; 137789 MW; 3BC6152134271F4D CRC64;
 Query Match 65.6%; Score 42; DB 1; Length 1453;
 Match Local Similarity 100.0%; Pred. No. 6.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 POGINGOR 9
 Db 940 POGINGOR 947

RESULT 3
 CALL_MOUSE STANDARD; PRT: 1453 AA.
 AC P11087; O60635;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN COL1A1 OR COLA1.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murineae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N;
 RX MEDLINE=96033240; PubMed=8535610;
 RA Li S.W., Khillan J., Prockop D.J.;
 RT "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
 of type I procollagen.";
 RL Matrix Biol. 14:593-595(1995).
 RN [2]
 RP SEQUENCE OF 518-1128 FROM N.A.
 RX MEDLINE=86137403; PubMed=3841523;
 RA French B.T., Lee W.-H., Maul G.G.;
 RT "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
 collagen protein.";
 RL Gene 39:311-312(1985).
 RN [3]
 RP SEQUENCE OF 735-1130 FROM N.A.

RX MEDLINE=83141374; PubMed=6298597;
 RA Monson J.M., Friedman J., McCarthy B.J.;
 RT "DNA sequence analysis of a mouse pro alpha 1(I) procollagen gene:
 evidence for a mouse B1 element within the gene.";
 RL Mol. Cell. Biol. 2:1362-1371(1982).
 RN [4]
 RP SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
 RX MEDLINE=83157109; PubMed=6219867;
 RA Monson J.M., McCarthy B.J.;
 RT "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
 evidence for insertions or deletions in gene coding sequences.";
 RL DNA 1:59-69(1981).
 RN [5]
 RP SEQUENCE OF 1442-1453 FROM N.A.
 RX MEDLINE=88124276; PubMed=3340560;
 RA Mooslehner K., Harbers K.;
 RT "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
 of the 3'-untranslated region.";
 RL Nucleic Acids Res. 16:773-773(1988).
 CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 (FIBRILLAR FORMING COLLAGEN).
 CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
 CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
 BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
 HYDROXYAPATITE.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- SIMILARITY: CONTAINS 1 VWC DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: 008020; AAA88912.1; -
 DR EMBL: X15896; CAA33904.1; -
 DR EMBL: M14423; AAA37333.1; -
 DR EMBL: M17491; AAA37334.1; -
 DR EMBL: X06753; CAA29927.1; -
 DR EMBL: K03036; AAA37332.1; -
 DR EMBL: K03028; AAA37332.1; JOINED.
 DR EMBL: K03030; AAA37332.1; JOINED.
 DR EMBL: K03031; AAA37332.1; JOINED.
 DR EMBL: K03032; AAA37332.1; JOINED.
 DR EMBL: K03033; AAA37332.1; JOINED.
 DR EMBL: K03034; AAA37332.1; JOINED.
 DR EMBL: K03035; AAA37332.1; JOINED.
 DR PIR: A23982; A23982.
 DR MGD; MGI:88467; Col1a1.
 DR InterPro: IPR000087; Collagen.
 DR InterPro: IPR000885; Fib_collagen_C.
 DR InterPro: IPR001007; VWF_C.
 DR Pfam: PF01391; Collagen; 18.
 DR Pfam: PF01410; COLFI; 1.
 DR ProDom: PD000007; Collagen; 1.
 DR ProDom: PD002078; Fib_collagen_C; 1.
 DR SMART: SM00038; COLFI; 1.
 DR SMART: SM00214; VWC; 1.
 DR PROSITE: PS01208; VWF; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 22
 FT PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
 FT CHAIN 152 1207 COLLAGEN ALPHA 1(I) CHAIN.
 FT PROPEP 1208 1453 CARBOXYL-TERMINAL PROPEPTIDE.
 FT DOMAIN 29 87 VWC.
 FT DOMAIN 152 167 NONHELICAL REGION (N-TERMINAL).
 FT DOMAIN 168 1181 TRIPLE-HELICAL REGION.
 FT DOMAIN 1182 1207 NONHELICAL REGION (C-TERMINAL).
 FT CARBOHYD 56 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1354 1354 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SITE 734 736 CELL ATTACHMENT SITE (POTENTIAL).
 FT CONFLICT 1082 1084 CELL ATTACHMENT SITE (POTENTIAL).
 FT SEQUENCE 1453 AA: 137944 MW: 38025535DF81808 CRC4:

Query Match
 Best Local Similarity 65.6%; Score 42; DB 1; Length 1453;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 POGIAGOR 9
 DB 940 POGIAGOR 947

RESULT 4
 ID CALL_CANFA STANDARD: PRT; 1460 AA.
 AC 09XSJ7:
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN COL1A1.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin;
 RA Campbell B.G., Woolton J.A.M., McLeod J.N., Minor R.R.;
 RT "Sequence of normal canine COL1A1 cDNA."
 RT Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
 (FIBRILLAR FORMING COLLAGEN).
 CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
 CC -1- PTM: POLYLINES AT THE THIRD POSITION OF THE TRIPLEPEPTIDE REPEATING
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF153062; AAD34619.1; .
 DR InterPro: IPR000087; Collagen.
 DR InterPro: IPR000885; Fib.collagen_C.
 DR InterPro: IPR001007; VWFC_C.
 DR Pfam: PF01391; Collagen; 18.
 DR Pfam: PF01410; COLFI; 1.
 DR ProDom: PD000007; Collagen; 1.
 DR ProDom: PD002078; Fib.collagen_C; 1.
 DR SMART: SM00214; VWFC; 1.
 DR SMART: PS01208; VWFC; 1.
 DR PROSITE: PS01208; VWFC; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KM Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 22
 FT PROPEP 23 157
 FT CHAIN 158 1214 AMINO-TERMINAL PROPEPTIDE.
 FT PROPEP 1215 1460 COLLAGEN ALPHA 1(I) CHAIN.
 FT DOMAIN 34 92 CARBOXYL-TERMINAL PROPEPTIDE.
 FT DOMAIN 158 174
 FT DOMAIN 175 1188 NONHELICAL REGION (N-TERMINAL).
 FT DOMAIN 1189 1214 TRIPLE-HELICAL REGION.
 FT SITE 741 743 NONHELICAL REGION (C-TERMINAL).
 FT SITE 1089 1091 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 1361 1361 CELL ATTACHMENT SITE (POTENTIAL).
 FT CARBOHYD 1361 1361 N-LINKED (GLCNAC. . .) (POTENTIAL).

SQL SEQUENCE 1460 AA; 138762 MW; 58E3674D28570697 CRC64;

Query Match
 Best Local Similarity 65.6%; Score 42; DB 1; Length 1460;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 POGIAGOR 9
 DB 947 POGIAGOR 954

RESULT 5
 ID CALL_HUMAN STANDARD: PRT; 1464 AA.
 AC P02452; 015176; 014037;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN COL1A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RP [1]
 RP SEQUENCE OF 1-472 FROM N.A.
 RC MEDLINE=89025644; PubMed=3178743;
 RA Tromp G., Kuivaniemi H., Stacey A., Shketa H., Baldwin C.T.,
 RA Jaenisch R., Prockop D.J.;
 RT "Structure of a full-length cDNA clone for the prepro alpha 1(I)
 chain of human type I procollagen."
 RT Biochem. J. 253:919-922(1988).
 RL [2]
 RL SEQUENCE OF 1-181 FROM N.A.
 RC MEDLINE=84270697; PubMed=6462220;
 RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
 RA Myers J., Williams C., Ramirez F.;
 RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
 conservation of a pattern of introns and exons."
 RT Nature 310:337-340(1984).
 RL [3]
 RL SEQUENCE OF 162-301.
 RC TISSUE=Skin;
 RA MEDLINE=71001508; PubMed=5529814;
 RA Click E.M., Bornstein P.;
 RT "Isolation and characterization of the cyanogen bromide peptides from
 the alpha 1 and alpha 2 chains of human skin collagen."
 RT Biochemistry 9:4699-4706(1970).
 RL [4]
 RL SEQUENCE OF 263-268.
 RC TISSUE=SKIN;
 RA MEDLINE=71001508; PubMed=4319110;
 RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
 RT "A comparative study of glycopeptides derived from selected
 vertebrate collagens. A possible role of the carbohydrate in fibril
 formation."
 RT J. Biol. Chem. 245:5042-5048(1970).
 RL [5]
 RL SEQUENCE OF 425-1464 FROM N.A.
 RC MEDLINE=84080385; PubMed=6689127;
 RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
 RA Prockop D.J.;
 RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
 pro alpha 1 chain of human type I procollagen. Statistical evaluation
 of structures that are conserved during evolution."
 RT Biochemistry 22:5213-5223(1983).
 RL [6]
 RL SEQUENCE OF 1229-1454 FROM N.A.
 RC TISSUE=BONE;
 RA MEDLINE=88124208; PubMed=3340531;
 RA Meekelae J.K., Raassina M., Virta A., Vuorio E.;
 RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
 domain."

DB 951 POCIACOR 958

|||||||

RESULT 6
ID CA12_HUMAN STANDARD; PRT; 1418 AA.
AC P02458;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE COL2A1 chain precursor [Contains: Chondrocalcin].
GN COL2A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RA MEDLINE=90067946; PubMed=2587267;
RA Su M.W., Lee B., Ramirez F., Machado M., Horton W.;
RT "Nucleotide sequence of the full length cDNA encoding for human type
RT II procollagen.";
RL Nucleic Acids Res. 17:9473-9473(1989).
RN 12
RP SEQUENCE OF 1-28 FROM N.A.
RA MEDLINE=87031574; PubMed=3021582;
RA Nunez A.M., Kohno K., Martin G.R., Yamada Y.;
RT "Promoter region of the human pro-alpha 1(II)-collagen gene.";
RL Gene 44:11-16(1986).
RN 13
RP SEQUENCE OF 432-1145 FROM N.A.
RA Ramirez F.;
RT Submitted (DEC-1988) to the EMBL/Genbank/DBJ databases.
RN 14
RP SEQUENCE OF 963-1418 FROM N.A.
RA MEDLINE=85190534; PubMed=3857598;
RA Chesh K.S.E., Stoker N.G., Griffin J.R., Grosveld F.G., Solomon E.;
RT "Identification and characterization of the human type II collagen
RT gene (COL2A1).";
RL Proc. Natl. Acad. Sci. U.S.A. 82:2555-2559(1985).
RN 15
RP SEQUENCE OF 1120-1398 FROM N.A.
RA MEDLINE=85306861; PubMed=3840017;
RA Elima K., Vuorio E.;
RT "Determination of the single polyadenylation site of the human pro
RT alpha 1(II) collagen gene.";
RL Nucleic Acids Res. 15:9499-9504(1987).
RN 17
RP SEQUENCE OF 1227-1289 FROM N.A.
RA MEDLINE=86104139; PubMed=3002437;
RA Nunez A.M., Francomano C., Young M.F., Martin G.R., Yamada Y.;
RT "Isolation and partial characterization of genomic clones coding for
RT a human pro-alpha 1 (II) collagen chain and demonstration of
RT restriction fragment length polymorphism at the 3' end of the gene.";
RL Biochemistry 24:6343-6348(1985).
RN 18
RP SEQUENCE OF 1176-1226 FROM N.A.
RA MEDLINE=8418798; PubMed=6320112;
RA Strom C.M., Upholt W.B.;
RT "Isolation and characterization of genomic clones corresponding to
RT the human type II procollagen gene.";
RL Nucleic Acids Res. 12:1025-1038(1984).
RN 19
RP SEQUENCE OF 35-167 FROM N.A.

RX MEDLINE=89233138; PubMed=2714801;
RA Su M.W., Benson-Chanda V., Vissing H., Ramirez F.;
RT "Organization of the exons coding for pro alpha 1(II) collagen N-
RT propeptide confirms a distinct evolutionary history of this domain of
RT the fibrillar collagen genes.";
RL Genomics 4:438-441(1989).
RN 10
RP REVIEW ON VARIANTS.
RA MEDLINE=91184577; PubMed=2010058;
RA Kuvantem H., Tromp G., Prockop D.J.;
RT "Mutations in collagen genes: causes of rare and some common diseases
RT in humans.";
RL FASEB J. 5:2052-2060(1991).
RN 11
RP REVIEW ON VARIANTS.
RA MEDLINE=97253593; PubMed=9101290;
RA Kuvantem H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
RN 12
RP VARIANT SER-1074.
RA MEDLINE=90036909; PubMed=2572591;
RA Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,
RA Hollister D.W.;
RT "Glycine to serine substitution in the triple helical domain of pro-
RT alpha 1(II) collagen results in a lethal perinatal form of short-
RT limbed dwarfism.";
RL J. Biol. Chem. 264:18265-18267(1989).
RN 13
RP VARIANT SDC-1095-GLY--TYR-1330 DEL.
RA MEDLINE=89266907; PubMed=2543071.
RA Lee B., Vissing H., Ramirez F., Rogers D., Rimoin D.;
RT "Identification of the molecular defect in a family with
RT spondyloepiphyseal dysplasia.";
RL Science 244:978-980(1989).
RN 14
RP VARIANT OSTEOARTHRITIS CVS-650.
RA MEDLINE=90370826; PubMed=1975693;
RA Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;
RT "Single base mutation in the type II procollagen gene (COL2A1) as a
RT cause of primary osteoarthritis associated with a mild
RT chondrodysplasia.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).
RN 15
RP VARIANT OF-IV VAL-717.
RA MEDLINE=91291136; PubMed=2064612;
RA Baleman J.F., Hannagan M., Chan D., Cole W.G.;
RT "Characterization of a type I collagen alpha 2(I) glycine-586 to
RT valine substitution in osteogenesis imperfecta type IV. Detection of
RT the mutation and prenatal diagnosis by a chemical cleavage method.";
RL Biochem. J. 276:765-770(1991).
RN 16
RP VARIANT OSTEOARTHRITIS CVS-650.
RA MEDLINE=91086471; PubMed=1985108;
RA Eyre D.R., Weis M.A., Moskowitz R.W.;
RT "Cartilage expression of a type II collagen mutation in an inherited
RT form of osteoarthritis associated with a mild chondrodysplasia.";
RL J. Clin. Invest. 87:357-361(1991).
RN 17
RP VARIANT HYPOCHONDROGENESIS GIU-984.
RA MEDLINE=93054548; PubMed=1429602;
RA Bogaert R., Tiller G.E., Wiles M.A., Gruber H.E., Rimoin D.L.,
RA Cohn D.H., Eyre D.R.;
RT "An amino acid substitution (Gly853-->Glu) in the collagen alpha
RT 1(II) chain produces hypochondrogenesis.";
RL J. Biol. Chem. 267:22522-22526(1992).
RN 18
RP VARIANT HYPOCHONDROGENESIS SER-705.
RA MEDLINE=92262484; PubMed=1374906;
RA Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,
RA Ramirez F., Vitale E., Lee B.;

RT "Characterization of a type II collagen gene (COL2A1) mutation
RT identified in cultured chondrocytes from human hypochondrogenesis.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).
RN [19]
RP VARIANT WS-II ASP-198.
RX MEDLINE=93304428; PubMed=83171498;
RA Koerkoe J., Rittvaient P., Haataja L., Kaeerlaeinen H.,
RA Kivirikko K.I., Prockop D.J., Ala-Kokko L.;
RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate
RT for glycine alpha 1-67 and that causes cataracts and retinal
RT detachment: evidence for molecular heterogeneity in the Wagner
RT syndrome and the Stickler syndrome (arthro-ophthalmopathy).";
RL Am. J. Hum. Genet. 53:55-61(1993).
RN [20]
RP VARIANT SEMD CVS-840.
RA Eyle D.R.; Weis M.A., Lachman R.S., Cohn D.H., Rimoia D.L.,
RT "A dominant mutation in the type II collagen gene (COL2A1) produces
RT spondyloepiphyseal dysplasia (SEMD), Strudwick type.";
RL Am. J. Hum. Genet. 53:A209-A209(1993).
RN [21]
RP VARIANT OSTEOARTHRITIS CVS-650.
RX MEDLINE=93282819; PubMed=8507190;
RA Holderbaum D., Malemud C.J., Moskowitz R.W., Haqgi T.M.;
RT "Human cartilage from late stage familial osteoarthritis transcribes
RT type II collagen mRNA encoding a cysteine in position 519.";
RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).
RN [22]
RP VARIANT SEMD ARG-285.
RX MEDLINE=93252400; PubMed=8486375;
RA Peltonen L., Rittvaient P., Vuorio A.F., Kaitila I., Ala-Kokko L.,
RT "A mutation in the amino-terminal end of the triple helix of type II
RT collagen causing severe osteochondrodysplasia.";
RL Genomics 16:282-285(1993).
RN [23]
RP VARIANT SEMD CVS-206.
RX MEDLINE=94063862; PubMed=8244341;
RA Williams C.J., Considine E.L., Knowlton R.G., Reginato A., Neumann G.,
RA Harrison D., Buxton P., Jimenez S.A., Prockop D.J.;
RT "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a
RT family with an Arg75->Cys mutation in the procollagen type II gene
RT (COL2A1).";
RL Hum. Genet. 92:499-505(1993).
RN [24]
RP VARIANT SEMD CVS-920.
RX MEDLINE=93315508; PubMed=8325895;
RA Chan D., Taylor T.K.F., Cole W.G.;
RT "Characterization of an arginine 789 to cysteine substitution in
RT alpha 1 (II) collagen chains of a patient with spondyloepiphyseal
RT dysplasia.";
RL J. Biol. Chem. 268:15238-15245(1993).
RN [25]
RP VARIANT SEMD SER-1128.
RX MEDLINE=93140139; PubMed=8423604;
RA Cole W.G., Hall R.K., Rogers J.G.;
RT "The clinical features of spondyloepiphyseal dysplasia congenita
RT resulting from the substitution of glycine 997 by serine in the alpha
RT 1(II) chain of type II collagen.";
RL J. Med. Genet. 30:27-35(1993).

Query Match 62.5%; Score 40; DB 1; Length 1418;
Best Local Similarity 87.5%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 POGIAGOR 9
DB 904 POGIAGOR 911
RESULT 7
CA12_MOUSE STANDARD; PRT; 1459 AA.
ID CA12_MOUSE

AC P28481;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(II) chain precursor [Contains: Chondrocalcin].
GN COL2A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91358489; PubMed=1885613;
RA Metasanta M., Toman D., de Crombrughe B., Vuorio E.;
RT "Mouse type II collagen gene. Complete nucleotide sequence, exon
RT structure, and alternative splicing";
RL J. Biol. Chem. 266:16862-16869(1991).
RN [2]
RP SEQUENCE OF 1455-1459 FROM N.A.
RX MEDLINE=91274355; PubMed=2054384;
RA Metasanta M., Toman D., de Crombrughe B., Vuorio E.;
RT "Specific hybridization probes for mouse type I, II, III and IX
RT collagen mRNAs.";
RL Biochim. Biophys. Acta 1089:241-243(1991).
CC -1- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.
CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.
CC -1- PTM: POLYMER AT THE THIRD POSITION OF THE TRIPLE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M65161; AAA68100.1; -
CC EMBL: X57982; CAA41047.1; -
CC MGD: MGI:88452; COL2A1.
DR InterPro: IPR000085; FIB_collagen_C.
DR InterPro: IPR000087; FIB_collagen_C.
DR InterPro: IPR001007; VWFC_C.
DR Pfam: PF00093; vwc; 1.
DR Pfam: PF01391; Collagen; 18.
DR Pfam: PF01410; COLF1; 1.
DR ProDom: PD000007; Collagen; 3.
DR ProDom: PD002078; FIB_collagen_C; 1.
DR SMART: SM00038; COLF1; 1.
DR SMART: SM00214; vwc; 1.
DR PROSITE: PS01208; VWFC; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Cartilage; Signal; Alternative splicing.
FT SIGNAL 1 25
FT PROPEP 26 153
FT CHAIN 154 1213
FT PROPEP 1214 1459
FT DOMAIN 32 89
FT DOMAIN 173 1186
FT DOMAIN 1187 1213
FT VARSPLIC 29 29
FT VARSPLIC 30 98
SQ SEQUENCE 1459 AA; 139154 MW; F6C84FA7C532E72 CRC64;
Query Match 62.5%; Score 40; DB 1; Length 1459;
Best Local Similarity 87.5%; Pred. No. 15;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 POGIAGOR 9
DB 904 POGIAGOR 911
RESULT 7
CA12_MOUSE STANDARD; PRT; 1459 AA.
ID CA12_MOUSE

DB 945 POGIAGOR 952

RESULT 8
PURA_NEIMA STANDARD: PRT; 1320 AA.

AC 09JMC5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phosphoribosylformylglycinamide synthase (EC 6.3.5.3) (FGAM synthase) (FGAMS) (Formylglycinamide ribotide synthetase)
DE (FGARAT) (Formylglycinamide ribotide synthetase)
GN PURL OR NMA0445.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
NCBI_TaxID=65599;

RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S., Rajendram M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrell B.G.;
RA "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491".
RT Nature 404:502-506(2000).
RL

CC -1- CATALYTIC ACTIVITY: ATP + N(2)-formyl-N(1)-(5-phospho-D-riboseyl)glycinamide + L-glutamine + H(2)O = ADP + phosphate + 2-(formamido)-N(1)-(5-phospho-D-riboseyl)acetamide + L-glutamate.
CC -1- PATHWAY: De novo purine biosynthesis; fourth step.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE FGAMS FAMILY.
CC -1- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC

DR EMBL; AL162753; CAB83743.1; -
DR InterPro; IPR000728; AIRS-related.
DR Pfam; PF00586; AIRS; 1.
DR Pfam; PF02769; AIRS_C; 2.
KW Purine biosynthesis; Ligase; ATP-binding; Glutamine amidotransferase;
KW Complete proteome.
FT NP_BIND 311 322 ATP (POTENTIAL).
FT ACT_SITE 1162 1162 GATASE (BY SIMILARITY).
SQ SEQUENCE 1320 AA; 143790 MW; FEB32DC315CEDD50 CRC64;

Query Match 60.9%; Score 39; DB 1; Length 1320;
Best Local Similarity 80.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 SPOGIAGORN 10
ID 1265 SPOGIAGVTN 1274

RESULT 9
PURA_NEIMA STANDARD: PRT; 1320 AA.
AC 09JMK5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phosphoribosylformylglycinamide synthase (EC 6.3.5.3) (FGAM synthase) (FGAMS) (Formylglycinamide ribotide synthetase)
DE (FGARAT) (Formylglycinamide ribotide synthetase)
GN PURL OR NMB1996.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
NCBI_TaxID=491;

RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwinn S.L., Deboy R., Peterson J.D., Hickey E.K., Halt D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Clecko A., Parksey D.S., Blair E., Citron H., Clark E.B., Cotton M.D., Uitterback T.R., Khouri H., Qin H., Vamathevan J., Gill J., Scarlato V., Maignani V., Pizze M., Grandi C., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappaport R., Venter J.C.;
RA "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58".
RT Science 287:1809-1815(2000).
RL

CC -1- CATALYTIC ACTIVITY: ATP + N(2)-formyl-N(1)-(5-phospho-D-riboseyl)glycinamide + L-glutamine + H(2)O = ADP + phosphate + 2-(formamido)-N(1)-(5-phospho-D-riboseyl)acetamide + L-glutamate.
CC -1- PATHWAY: De novo purine biosynthesis; fourth step.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE FGAMS FAMILY.
CC -1- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC

DR EMBL; AE002549; AAF42323.1; -
DR TIGR; NMB1996; -
DR InterPro; IPR000728; AIRS-related.
DR Pfam; PF00586; AIRS; 1.
DR Pfam; PF02769; AIRS_C; 2.
KW Purine biosynthesis; Ligase; ATP-binding; Glutamine amidotransferase;
KW Complete proteome.
FT NP_BIND 311 322 ATP (POTENTIAL).
FT ACT_SITE 1162 1162 GATASE (BY SIMILARITY).
SQ SEQUENCE 1320 AA; 143852 MW; B6B873913EBB06BD CRC64;

Query Match 60.9%; Score 39; DB 1; Length 1320;
Best Local Similarity 80.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 SPOGIAGORN 10
ID 1265 SPOGIAGVTN 1274

RESULT 10
SELB_HAEIN STANDARD: PRT; 619 AA.
AC P43927;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Selenocysteine-specific elongation factor (SelB translation factor).
GN SELB OR H10709.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae; Haemophilus.

NCBI_TaxID=727;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-T., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedlorn E., Cotton M.D.,
 RA Uutterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fritchman J.L., Geoghagen N.S.M.,
 RA Gheue C.B., McDonald J.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 Rd.";
 RL Science 269:496-512(1995).
 CC -1- FUNCTION: TRANSLATION FACTOR NECESSARY FOR THE INCORPORATION OF
 CC SELENOCYSTEINE INTO PROTEINS. IT PROBABLY REPLACES EF-TU FOR THE
 CC INSERTION OF SELENOCYSTEINE DIRECTED BY THE UGA CODON. SELB BINDS
 CC GTP AND GDP (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
 CC SELB SUBFAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: U32753; AAC22366.1; -
 DR HSSP: P02990; 1EFU.
 DR TIGR: H10709; -
 DR InterPro: IPR000591; DEP.
 DR InterPro: IPR004161; EFTU_D2.
 DR InterPro: IPR000795; EF_GTPbind.
 DR InterPro: IPR005225; Small_GTP.
 DR InterPro: IPR004535; TEF_SelB.
 DR Pfam: PF00009; GTP_EFTU; 1.
 DR Pfam: PF00610; DEP; 1.
 DR Pfam: PF03144; GTP_EFTU_D2; 1.
 DR TIGRfams: TIGR00231; small_GTP; 1.
 DR TIGRfams: TIGR00475; selb; 1.
 DR PROSITE: PS00301; EFATOR_GTP; FALSE_NEG.
 KM Protein biosynthesis: GTP-binding; Complete proteome.
 FT NP_BIND 7 14 GTP (BY SIMILARITY).
 FT NP_BIND 56 60 GTP (BY SIMILARITY).
 FT NP_BIND 111 114 GTP (BY SIMILARITY).
 SQ SEQUENCE 619 AA: 69975 MW: 5960BACDE578D404 CRC64;
 Query Match 60.2%; Score 38.5; DB 1; Length 619;
 Best Local Similarity 66.7%; Pred. No. 12;
 Matches 10; Conservative 0; Mismatches 2; Indels 3; Gaps 1;
 QY 1 SPOGIAGOR---NPN 12
 Db 223 SEOGIAGORLALN 237
 RESULT 11
 GAG_VILVK STANDARD; PRT; 442 AA.
 ID GAG_VILVK PRT; 442 AA.
 AC P35955;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE GAG polypeptide [contains: Core proteins P16, P25, P14].
 OS Vigna lentivirus (strain KVI772).
 GN GAG.
 OC Virenses; Retrovirus; Retroviridae; Lentivirus.

NCBI_TaxID=36374;
 [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=93174981; PubMed=8382414;
 RA Anderson O.S., Eiser J.E., Tobin G.J., Greenwood J.D., Gonda M.A.,
 RA Georgsson G., Andersdotter V., Benediktsson E., Carlsson H.M.,
 RA Maenylla E.O., Rafnar B., Palsson P.A., Casey J.W., Petursson G.;
 RT "Nucleotide sequence and biological properties of a pathogenic
 RT porcine molecular clone of neuroinfectious virus.";
 RL Virology 193:89-105(1993).
 CC -1- FUNCTION: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- SIMILARITY: CONTAINS 2 CCHC-TYPE ZINC FINGERS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: L06906; AAA48358.1; -
 DR EMBL: S55323; AAB25459.1; -
 DR PIR: A45390; A45390.
 DR InterPro: IPR00721; Gag_P24.
 DR InterPro: IPR01878; ZnF_CCHC.
 DR Pfam: PF00098; zf-CCHC; 2.
 DR Pfam: PF00607; Gag_P24; 1.
 DR PRINTS: PR00939; C2HCZNFINGER.
 DR SMART: SM00343; ZnF_C2HC; 2.
 DR PROSITE: PS50158; zf_CCHC; 2.
 KW Core protein; Polyprotein; Zinc-finger; Repeat.
 FT CHAIN 1 143 CORE PROTEIN P16.
 FT CHAIN 144 363 CORE PROTEIN P25.
 FT CHAIN 364 442 CORE PROTEIN P14.
 FT ZN_FING 385 402 CCHC-TYPE 1.
 FT ZN_FING 404 421 CCHC-TYPE 2.
 SQ SEQUENCE 442 AA: 49856 MW: 55EB18E951B486FA CRC64;
 Query Match 57.8%; Score 37; DB 1; Length 442;
 Best Local Similarity 63.6%; Pred. No. 16;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 2 POGIAGORNNFN 12
 Db 374 POGKAGQKGVN 384
 RESULT 12
 BISC_ECOLI STANDARD; PRT; 777 AA.
 ID BISC_ECOLI PRT; 777 AA.
 AC P20099;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Biotin sulfoxide reductase (Ec 1.-.-.-) (BDS reductase) (BSO
 DE reductase).
 GN BISC OR B3551.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=90202748; PubMed=2180922;
 RA Pierson D.E., Campbell A.;
 RT "Cloning and nucleotide sequence of bisc, the structural gene for
 RT biotin sulfoxide reductase in Escherichia coli.";
 RL J. Bacteriol. 172:2194-2198(1990).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=94316500; PubMed=8041620;

```

RT proteol. Papd of Escherichia coli. ";
RT FEBS Lett. 286:79-82(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-KIM5; PLASMid-pMT1 (pMT-1);
RX MEDLINE=99043898; PubMed=9826348;
RA Lindler L.E., Plano G.V., Burland V., Mayhew G.F., Blattner F.R.;
RT "Complete DNA sequence and detailed analysis of the Yersinia pestis
RL KIM5 plasmid encoding murine toxin and capsular antigen. ";
RN Infect. Immun. 66:5731-5742(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-KIM5; PLASMid-pMT1 (pMT-1);
RX MEDLINE=98422474; PubMed=9748454;
RA Hu P., Elliott J., McCreedy P., Showronski E., Barnes J.,
RT "Structural organization of virulence-associated plasmids of Yersinia
RL pestis. ";
RN J. Bacteriol. 180:5192-5202(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-CO-92 / Biovar Orientalis; PLASMid-pMT1 (pMT-1);
RX MEDLINE=21470413; PubMed=11586360;
RA Parthall J., Wren B.W., Thomson N.R., Tlball R.W., Holden M.T.G.,
RA Prentice M.B., Sebathila M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague. ";
RN Nature 413:523-527(2001).
RC CC -1- FUNCTION: HAS A STIMULATORY ROLE FOR THE ENVELOPE ANTIGEN E1
CC SECRETION. IT SEEMS TO INTERACT WITH THE SUBUNIT POLYPEPTIDE AND
CC TO PREVENT IT FROM DIGESTION BY A PROTEASE.
CC CC -1- SIMILARITY: BELONGS TO THE PERIPLASMIC PILUS CHAPERONE FAMILY.
CC CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE DOMAIN.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X61996; CAA43967.1; -.
DR EMBL: AF074611; AAC82756.1; ALT_INIT.
DR EMBL: AF053947; AAC13220.1; -.
DR EMBL: AL117211; CAB55264.1; -.
DR PIR: S16965; S16965.
DR PIR: S19096; S19096.
DR HSSP: P31697; 1BRF.
DR InterPro: IPR001829; Pili_chaperone.
DR Pfam: PF00345; Pili_assembly.2.
DR Pfam: PF02753; Pili_assembly.C; 2.
DR PRINTS: PR00969; CHAPERONPILI.
DR ProDom: PD001447; Pili_chaperone; 1.
DR PROSITE: PS00635; PILI_CHAPERONE; 1.
KW Chaperone; Periplasmic; Signal; Immunoglobulin domain; Plasmid;
KW complete proteome.
-FT SIGNAL 1 20 POTENTIAL.
-FT CHAIN 21 258 CHAPERONE PROTEIN CAF1M.
-FT DISULEFID 121 160 POTENTIAL.
-FT CONFLICT 180 180 N -> K (IN REF. 1).
SQ SEQUENCE 258 AA; 28751 MW; 8049DCC1A80C7391 CRC64;

```

Db 228 PKGLAGARN 236

11:11 11

RESULT 14
MATT_MARGL STANDARD; PRT; 324 AA.
ID MATT_MARGL
AC P51950;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cdk-activating kinase assembly factor MATT (RING finger protein MATT)
DE (Mange a trois) (CDK/cyclin H assembly factor).
OS *Marthasterias glacialis* (Splyn starfish).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Asterozoa;
OC Asteroidea; Forcipulatacea; Forcipulatida; Asteroidea; *Marthasterias*.
OX NCBI_TaxID=7609;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Oocyte;
RA MEDLINE=96067131; PubMed=7588631;
RA Devault A., Martinez A.-M., Fesquet D., Labbe J.-C., Morin N.,
RA Tassan J.-P., Nigg E.A., Cavadore J.-C., Doree M.,
RT "MATT" (Mange a trois'), a new RING finger protein subunit
RT stabilizing cyclin H-CDK7 complexes in starfish and *Xenopus* CAK.";
RL EMBL J. 14:5027-5036(1995).
CC -1- FUNCTION: STABILIZES THE CYCLIN H-CDK7 COMPLEX TO FORM A
CC -1- FUNCTIONAL CDK-ACTIVATING KINASE (CAK) ENZYMACTIC COMPLEX.
CC -1- SUBUNIT: ASSOCIATES WITH CDK7 AND CYCLIN H.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U29666; AAC46933.1; -
DR InterPro: IPR004575; Cdk7.
DR InterPro: IPR004599; Tfb3.
DR InterPro: IPR001841; Znf_ring.
DR Pfam: PF00097; Zf-C3HC4; 1.
DR SMART: SM00184; RING; 1.
DR TIGRFAMs: TIGR00570; cdk7; 1.
DR TIGRFAMs: TIGR00626; tfb3; 1.
DR PROSITE: PS00518; ZF_RING_1; 1.
DR PROSITE: PS00589; ZF_RING_2; 1.
DR Cell cycle, cell division, Nuclear protein; zinc-finger.
FT ZNFING 7 51 RING-TYPE.
SQ SEQUENCE 324 AA; 37432 MW; 69FBDFFC7502C27F1 CRC64;
Query Match 56.2%; Score 36; DB 1; Length 324;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SPOGIAG 7
DB 280 SPOGIAG 286

OS *Solanum tuberosum* (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; *Solanum*.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Dianella;
RX MEDLINE=94105324; PubMed=8278528;
RA Poulsen P., Kreiberg J.D.;
RT "Starch branching enzyme cDNA from *Solanum tuberosum*.";
RL Plant Physiol. 102:1053-1054(1993).
RN [2]
RP SEQUENCE OF 279-527 FROM N.A.
RC STRAIN=cv. Desiree; TISSUE=tuber;
RX MEDLINE=92079917; PubMed=1745241;
RA Kossman J., Visser R.G.F., Mueller-Roeber B., Willmitzer L.,
RA Sonnewald U.;
RT "Cloning and expression analysis of a potato cDNA that encodes
RT branching enzyme: evidence for co-expression of starch biosynthetic
RT genes.";
RL Mol. Gen. Genet. 230:39-44(1991).
CC -1- CATALYTIC ACTIVITY: Formation of 1,6-glucosidic linkages of
CC starch.
CC -1- PATHWAY: Starch biosynthesis; third step.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Chloroplast; amyloplast.
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X69805; CAA9463.1; -
DR PIR: S18594; S18594.
DR InterPro: IPR000461; Alpha-amylase.
DR InterPro: IPR004193; Isoamylase_N.
DR Pfam: PF00128; alpha-amylase_N.
DR Pfam: PF02922; Isoamylase_N; 1.
KW Starch biosynthesis; transferase; Glycosyltransferase; Amyloplast.
FT ACT_SITE 424 424 BY SIMILARITY.
FT ACT_SITE 484 484 BY SIMILARITY.
FT ACT_SITE 553 553 BY SIMILARITY.
SQ SEQUENCE 861 AA; 99083 MW; F3D519AC7CF1BEF2 CRC64;
Query Match 56.2%; Score 36; DB 1; Length 861;
Best Local Similarity 57.1%; Pred. No. 50;
Matches 8; Conservative 2; Mismatches 2; Indels 2; Gaps 1;
QY 1 SPOGIAG--QRNFN 12
DB 744 SPOGIAGVETNFN 757

RESULT 16
ANDR_HUMAN STANDARD; PRT; 919 AA.
ID ANDR_HUMAN
AC P10275;
DT 01-MAR-1988 (Rel. 10, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Androgen receptor (Dihydrotestosterone receptor).
GN AR OR NR3C4 OR DHTT.
OS *Homo sapiens* (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

- RP SEQUENCE FROM N.A.
 RA MEDLINE=89112208; Pubmed=3216866;
 RA Lubahn D.B., Joseph D.R., Sar M., Tan J., Higgs H.N., Larson R.E.,
 RA French F.S., Wilson E.M.;
 RT "The human androgen receptor: complementary deoxyribonucleic acid
 RT cloning, sequence analysis and gene expression in prostate.";
 RL Mol. Endocrinol. 2:1265-1275(1988).
 RN [12]
 RP SEQUENCE FROM N.A., AND VARIANT CAIS MET-866.
 RA MEDLINE=90083302; Pubmed=2594783;
 RA Lubahn D.B., Brown T.R., Simental J.A., Higgs H.N., Migeon C.J.,
 RA Wilson E.M., French F.S.;
 RT "Sequence of the Intron/exon junctions of the coding region of the
 RT human androgen receptor gene and identification of a point mutation
 RT in a family with complete androgen insensitivity.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:9534-9538(1989).
 RN [13]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=90258935; Pubmed=2342476;
 RA Govindan M.V.;
 RT "Specific region in hormone binding domain is essential for hormone
 RT binding and trans-activation by human androgen receptor.";
 RL Mol. Endocrinol. 4:417-427(1990).
 RN [14]
 RP SEQUENCE FROM N.A.
 RA TISSUE=Prostate; Pubmed=3174628;
 RA MEDLINE=89017168; Pubmed=3174628;
 RA Chang C., Kokontis J., Liao S.;
 RT "Structural analysis of complementary DNA and amino acid sequences of
 RT human and rat androgen receptors.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:7211-7215(1988).
 RN [15]
 RP SEQUENCE FROM N.A.
 RA TISSUE=Prostate; Pubmed=2911578;
 RA MEDLINE=89098909; Pubmed=2911578;
 RA Tilley W.D., Marcelli M., Wilson J.D., McPhaul M.J.;
 RT "Characterization and expression of a cDNA encoding the human androgen
 RT receptor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:327-331(1989).
 RN [16]
 RP SEQUENCE FROM N.A.
 RA TISSUE=Prostate; Pubmed=2293020;
 RA MEDLINE=91155943; Pubmed=2293020;
 RA Marcelli M., Tilley W.D., Wilson C.M., Griffitt J.E., Wilson J.D.,
 RA McPhaul M.J.;
 RT "Definition of the human androgen receptor gene structure permits the
 RT identification of mutations that cause androgen resistance: premature
 RT termination of the receptor protein at amino acid residue 588 causes
 RT complete androgen resistance.";
 RL Mol. Endocrinol. 4:1105-1116(1990).
 RN [17]
 RP SEQUENCE OF 189-919 FROM N.A.
 RA MEDLINE=88178111; Pubmed=3353726;
 RA Chang C., Kokontis J., Liao S.;
 RT "Molecular cloning of human and rat complementary DNA encoding
 RT androgen receptors.";
 RL Science 240:324-326(1988).
 RN [18]
 RP SEQUENCE OF 468-919 FROM N.A.
 RA MEDLINE=88240407; Pubmed=3377788;
 RA Trepan J., Klaassen P., Kuiper G.G.J.M., van der Korput J.A.G.M.,
 RA Faber P.W., van Rooij H.C.J., Geurts van Kessel A., Voorhorst W.M.,
 RA Mulder E., Brinkmann A.O.;
 RT "Cloning, structure and expression of a cDNA encoding the human
 RT androgen receptor.";
 RL Biochem. Biophys. Res. Commun. 153:241-248(1988).
 RN [19]
 RP INTERACTION WITH RAN.
 RA MEDLINE=99329028; Pubmed=10400640;
 RA Hsiao P.-W., Lin D.-L., Nakao R., Chang C.;
 RT "The linkage of Kennedy's neuron disease to ARA24, the first
 RT identified androgen receptor polyglutamine region-associated
 RT coactivator.";
- RL J. Biol. Chem. 274:20229-20234(1999).
 RN [10]
 RP POLYMORPHISM OF POLY-GLN REGION.
 RA MEDLINE=92220629; Pubmed=1561105;
 RA Sledzews H.F.B.M., Oostra B.A., Brinkmann A.O., Trapman J.;
 RT "Trinucleotide repeat polymorphism in the androgen receptor gene
 RT (AR).";
 RL Nucleic Acids Res. 20:1427-1427(1992).
 RN [11]
 RP POLYMORPHISM OF POLY-GLY REGION.
 RA TISSUE=Blood;
 RA Lu J., Danielson M.;
 RT Submitted (Feb-1995) to the EMBL/Genbank/DBJ databases.
 RN [12]
 RP POLYMORPHISM OF POLY-GLN REGION.
 RA MEDLINE=97250535; Pubmed=9096391;
 RA Giovannucci E., Stampfer M.J., Krithivas K., Brown M., Brufsky A.,
 RA Talcott J., Hennekens C.H., Kantoff P.W.;
 RT "The CAG repeat within the androgen receptor gene and its
 RT relationship to prostate cancer.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:3320-3323(1997).
 RN [13]
 RP ERRATUM.
 RA Giovannucci E., Stampfer M.J., Krithivas K., Brown M., Dahl D.,
 RA Brufsky A., Talcott J., Hennekens C.H., Kantoff P.W.;
 RL Proc. Natl. Acad. Sci. U.S.A. 94:8272-8272(1997).
 RN [14]
 RP REVIEW ON VARIANTS.
 RA MEDLINE=93092459; Pubmed=1458719;
 RA Pinsky L., Trifiro M.A., Kaufman M., Beitel L.K., Whahre A.,
 RA Kazemi-Esfarjani P., Sabbaghian N., Lumbruso R., Alvarado C.,
 RA Vasilou M., Gottlieb B.;
 RT "Androgen resistance due to mutation of the androgen receptor.";
 RL Clin. Invest. Med. 15:456-472(1992).
 RN [15]
 RP REVIEW ON VARIANTS AIS.
 RA MEDLINE=93339360; Pubmed=8339746;
 RA Brown T.R., Scherer P.A., Chang Y.-T., Migeon C.J., Chirrt P.,
 RA Muroto K., Zhou Z.;
 RT "Molecular genetics of human androgen insensitivity.";
 RL Eur. J. Pediatr. 152 Suppl. 2:S62-S69(1993).
 RN [16]
 RP REVIEW ON VARIANTS.
 RA MEDLINE=94059770; Pubmed=8240973;
 RA Sultan C., Lumbruso S., Poujol N., Belon C., Boudon C.,
 RA Lombacero J.-M.;
 RT "Mutations of androgen receptor gene in androgen insensitivity
 RT syndromes.";
 RL J. Steroid Biochem. Mol. Biol. 46:519-530(1993).
 RN [17]
 RP REVIEW ON VARIANTS.
 RA MEDLINE=95023089; Pubmed=7937057;
 RA Patterson M.N., Hughes I.A., Gottlieb B., Pinsky L.;
 RT "The androgen receptor gene mutations database.";
 RL Nucleic Acids Res. 22:3560-3562(1994).
 RN [18]
 RP REVIEW ON VARIANTS.
 RA MEDLINE=95352489; Pubmed=7626493;
 RA Brinkmann A.O., Jenster G., Ris-Stalpers C., van der Korput J.A.G.M.,
 RA Bruggenwirth H.T., Boelmer A.L.M., Trapman J.;
 RT "Androgen receptor mutations.";
 RL J. Steroid Biochem. Mol. Biol. 53:443-448(1995).
 RN [19]
 RP REVIEW ON VARIANTS.
 RA MEDLINE=97169385; Pubmed=9016528;
 RA Gottlieb B., Trifiro M.A., Lumbruso R., Vasilou D.M., Pinsky L.;
 RT "The androgen receptor gene mutations database.";
 RL Nucleic Acids Res. 25:158-162(1997).
 RN [20]
 RP VARIANT LNCAP ALA-877.
 RA MEDLINE=91083633; Pubmed=2260966;
 RA Veldschoote J., Ris-Stalpers C., Kuiper G.G.J.M., Jenster G.,
 RA Berrevoets C.A., Klaassen E., van Rooij H.C.J., Trapman J.,

RA Brinkmann A.O., Mulder E.;
 RT "A mutation in the ligand binding domain of the androgen receptor of
 human LNCaP cells affects steroid binding characteristics and
 RT response to anti-androgens.";
 RL Biochem. Biophys. Res. Commun. 173:534-540(1990).
 RN [21]
 RP VARIANTS CAIS CYS-774; GLN-831 AND MET-866.
 RX MEDLINE-91186983; PubMed-2082179;
 RA Brown T.R., Lubahn D.B., Wilson E.M., French F.S., Migeon C.J.,
 RA Corfen J.L.;
 RT "Functional characterization of naturally occurring mutant androgen
 RT receptors from subjects with complete androgen insensitivity.";
 RL Mol. Endocrinol. 4:1759-1772(1990).
 RN [22]
 RP VARIANTS CYS-774.
 RX MEDLINE-91130758; PubMed-1856263;
 RA Marcelin M., Tilley W.D., Zoppi S., Griffin J.E., Wilson J.D.,
 RA McPhaul M.J.;
 RT "Androgen resistance associated with a mutation of the androgen
 RT receptor at amino acid 772 (Arg->Cys) results from a combination of
 RT decreased messenger ribonucleic acid levels and impairment of
 RT receptor function.";
 RL J. Clin. Endocrinol. Metab. 73:318-325(1991).
 RN [23]
 RP VARIANTS CAIS PRO-617.
 RX MEDLINE-91154385; PubMed-1999491;
 RA Marcelin M., Zoppi S., Grino P.B., Griffin J.E., Wilson J.D.,
 RA McPhaul M.J.;
 RT "A mutation in the DNA-binding domain of the androgen receptor gene
 RT causes complete testicular feminization in a patient with
 RT receptor-positive androgen resistance.";
 RL J. Clin. Invest. 87:1123-1126(1991).
 RN [24]
 RP VARIANTS PAIS CYS-763.
 RX MEDLINE-91185626; PubMed-2010552;
 RA McPhaul M.J., Marcelin M., Tilley W.D., Griffin J.E.,
 RA Isidro-Gutierrez R.F., Wilson J.D.;
 RT "Molecular basis of androgen resistance in a family with a qualitative
 RT abnormality of the androgen receptor and responsive to high-dose
 RT androgen therapy.";
 RL J. Clin. Invest. 87:1413-1421(1991).
 RN [25]
 Query Match 56.2%; Score 36; DB 1; Length 919;
 Best Local Similarity 66.7%; Pred. No. 53;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 OY 2 POGIAGORN 10
 DB 486 POGIAGORN 494

RT "The adenosine2 gene of Drosophila melanogaster encodes a
 RT formylglycinamide ribotide amidotransferase.";
 RL Genome 36:924-934(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Berkeley;
 RX MEDLINE-20196006; PubMed-10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer J.G., Chumpe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Plannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck P., Brockstein P., Brothier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Kocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalili D., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei M., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mervinov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reibert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskis R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -I- SIMILARITY: 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
 CC -I- SIMILARITY: CONTAINS 1 TYPE-1 GLUTAMINE AMIDOTRANSFERASE DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 DR EMBL: U00683; A046468.1;
 DR EMBL: AE003612; AAF5329.1;
 DR Flybase: FBgn0000052; ade2.
 DR InterPro: IPR000728; AIRS-related.
 DR Pfam: PF00586; AIRS; 1.
 DR Pfam: PF02769; AIRS; 2.
 DR Purine biosynthesis; Ligase; ATP-binding; Glutamine amidotransferase.
 FT NP_BIND 327 338
 FT ACT_SITE 1180 1180
 FT CONFLICT 558 558
 FT CONFLICT 625 625
 FT SEQUENCE 1354 AA; 148084 MW; 8AA167C2D2920B2C CRC64;

Query Match 56.2%; Score 36; DB 1; Length 1354;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SPOGING 7
 DB 1290 SPOGING 1296

RESULT 18
 CAL4_DROME STANDARD; PRT; 1775 AA.
 ID CAL4_DROME
 AC P08120;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Collagen alpha 1(IV) chain precursor.
 GN CG25C OR Dcgl.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OK NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89054012; PubMed=3142875;
 RA Blumberg B., Mackrell A.J., Fessler J.H.;
 RT "Drosophila basement membrane procollagen alpha 1(IV). II. Complete
 RT cDNA sequence, genomic structure, and general implications for
 RT supramolecular assemblies.";
 RL J. Biol. Chem. 263:18328-18337(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RL Blumberg B.;
 RA Theists (1987), University of California / Los Angeles, U.S.A.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Mackrell A.J.;
 RL Theists (1992), University of California / Los Angeles, U.S.A.
 RN [4]
 RP SEQUENCE OF 1065-1775 FROM N.A.
 RX MEDLINE=87194801; PubMed=3106346;
 RA Blumberg B., Mackrell A.J., Olson P.F., Kurkinen M., Monson J.M.,
 RA Natzie J.E., Fessler J.H.;
 RT "Basement membrane procollagen IV and its specialized carboxyl domain
 RT are conserved in Drosophila, mouse, and human.";
 RL J. Biol. Chem. 262:5947-5950(1987).
 RN [5]
 RP SEQUENCE OF 1355-1775 FROM N.A.
 RX MEDLINE=87246644; PubMed=3109906;
 RA Cecchini J.P., Knibbeher B., Mirre C., le Parco Y.;
 RT "Evidence for a type-IV-related collagen in Drosophila melanogaster.
 RT Evolutionary constancy of the carboxyl-terminal noncollagenous
 RT domain.";
 RL Eur. J. Biochem. 165:587-593(1987).
 RN [6]
 RP SEQUENCE OF 762-1230 FROM N.A.
 RX MEDLINE=82197577; PubMed=6210912;
 RA Monson J.M., Natzie J., Friedman J., McCarthy B.J.;
 RT "Expression and novel structure of a collagen gene in Drosophila.";
 RL Proc. Natl. Acad. Sci. U.S.A. 79:1761-1765(1982).
 CC -1- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.
 CC -1- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.
 CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH
 CC INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NCI
 CC DOMAINS.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NCI) AT THEIR C-TERMINUS. FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPLEPEPTIDE REPEATING

UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 THESE, LOCATED IN THE NCI DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 IV COLLAGENS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC -----
 CC EMBL; M23704; AAA28404.1; -;
 CC EMBL; M96575; AAB59184.1; -;
 CC EMBL; J02727; AAA28423.1; -;
 CC EMBL; M28334; AAA28422.1; -;
 CC EMBL; V00200; CAA23486.2; -;
 CC PIR; A31893; A31893;
 CC DR FLYBASE; FBgn0000299; Cg25C.
 CC DR InterPro; IPR000087; Collagen.
 CC DR InterPro; IPR001442; Procollagnc4.
 CC DR Pfam; PF01391; Collagen; 25.
 CC DR Pfam; PF01413; C4; 2.
 CC DR PRODOM; PD003923; Procollagnc4; 2.
 CC DR SMART; SM00111; C4; 2.
 CC KW Extracellular matrix; Connective tissue; Basement membrane;
 KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 23
 FT PROPEP 24 7
 FT CHAIN 1 1775
 FT DOMAIN ? 1544
 FT DISULFID 1545 1775
 FT DISULFID 1569 1655
 FT DISULFID 1599 1652
 FT DISULFID 1611 1617
 FT DISULFID 1674 1770
 FT DISULFID 1708 1767
 FT DISULFID 1720 1727
 FT CARBOHYD 72 72
 FT CONFLICT 948 948
 FT CONFLICT 997 997
 FT CONFLICT 1357 1357
 FT CONFLICT 1360 1360
 FT CONFLICT 1373 1373
 FT CONFLICT 1496 1496
 FT CONFLICT 1507 1511
 FT CONFLICT 1529 1529
 FT CONFLICT 1733 1733
 FT SEQUENCE 1775 AA; 174119 MW; 2DE5AB23149525CD CRC64;
 OY 2 POGIAGORNFN 12
 DB 1375 POGIAGORNFN 1385
 Query Match 56.2%; Score 36; DB 1; Length 1775;
 Best Local Similarity 54.5%; Pred. No. 1e-02;
 Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

RESULT 19
 UBP2_MOUSE STANDARD; PRT; 353 AA.
 ID UBP2_MOUSE
 AC O88623;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ubiquitin carboxyl-terminal hydrolase 2 (EC 3.1.2.15) (Ubiquitin
 DE thioesterase 2) (Ubiquitin-specific processing protease 2)
 DE (Deubiquitinating enzyme 2) (41 kDa ubiquitin-specific protease).
 GN UBP2 OR UBP41.
 OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RA Gong L., Yeh E.T.H.;
 RT "Cloning and expression of the human and mouse UBP41.";
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
 CC ubiquitin + a thiol.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19.
 CC -----
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF079565; AAC28393.1;
 DR MEROPS: C19.013;
 DR MGD; MGI:1858176; Usp2.
 DR InterPro: IPR001394; UCH-2.
 DR Pfam: PF00442; UCH-1; 1.
 DR Pfam: PF00443; UCH-2; 1.
 DR PROSITE: PS00972; UCH_2_1; 1.
 DR PROSITE: PS00973; UCH_2_2; 1.
 DR PROSITE: PS0235; UCH_2_3; 1.
 DR Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
 FT ACT_SITE 24 24 BY SIMILARITY.
 FT ACT_SITE 297 297 BY SIMILARITY.
 FT ACT_SITE 305 305 BY SIMILARITY.
 SO SEQUENCE 353 AA; 40581 MW; 4FFB39A225FE8F1 CRC64;
 QY 1 SPOGIAGORN 10 54.7% Score 35; DB 1; Length 353;
 Db 10 SAQGIAGLRN 19 70.0% Pred. No. 31;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 RESULT 20
 ECFE_YERPE STANDARD; PRT; 451 AA.
 AC Q8BHS9;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protease ecfE (EC 3.4.24.-).
 GN ECFE OR YPO1051.
 OS Yersinia pestis.
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Yersinia.
 OX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPRAIN-CO-92 / Biovar Orientalis;
 RA MEDLINE-21470413; PubMed-11586360;
 RA Parkhill J., Wren B.W., Thomson N.R., Tittall R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,
 RA Leather S., Monte S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.,
 RT "Genome sequence of Yersinia pestis, the causative agent of plague."
 RL Nature 413:523-527(2001).
 CC -1- FUNCTION: Degrades both heat shock sigma factors rpoE and rpoH (By
 CC similarity).

CC -1- COFACTOR: Zinc (Probable).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
 CC (By similarity).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M50B.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC -----
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ414146; CAC8893.1;
 DR MEROPS: M50.004;
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001193; Peptidase_M50.
 DR InterPro: IPR004387; Zn_Mprotease.
 DR InterPro: IPR000130; Zn_Mpeptase.
 DR Pfam: PF00595; PDZ; 1.
 DR Pfam: PF02163; Peptidase_M50; 1.
 DR SMART: SM00228; PDZ; 2.
 DR TIGRfam: TIGR00054; mem_zinc_metalprot; 1.
 DR PROSITE: PS0106; PDZ; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 DR Hydrolyase; Metalloprotease; Zinc; Transmembrane; Inner membrane;
 KW Complete proteome.
 FT METAL 22 22 ZINC (CATALYTIC) (POTENTIAL).
 FT ACT_SITE 23 23 POTENTIAL.
 FT METAL 26 26 ZINC (CATALYTIC) (POTENTIAL).
 FT TRANSMEM 98 120 POTENTIAL.
 FT TRANSMEM 377 399 POTENTIAL.
 FT TRANSMEM 427 446 POTENTIAL.
 FT DOMAIN 199 280 PDZ.
 SO SEQUENCE 451 AA; 49428 MW; 9EBE92728D25B8A8 CRC64;
 QY 1 SPOGIAGORNFN 12 54.7% Score 35; DB 1; Length 451;
 Db 135 SPOSIAGQANIS 146 58.3% Pred. No. 40;
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 RESULT 21
 UBP2_HUMAN STANDARD; PRT; 605 AA.
 AC Q75604; Q9BQ21;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Ubiquitin carboxyl-terminal hydrolase 2 (EC 3.1.2.15) (Ubiquitin
 DE thiolesterase 2) (Ubiquitin-specific processing protease 2)
 DE (Deubiquitinating enzyme 2) (41 kDa ubiquitin-specific protease).
 GN USP2 OR UBP41.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Placenta;
 RA Gong L., Yeh E.T.H.;
 RT "Cloning and expression of the human and mouse UBP41.";
 RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Lymph;
 RA Strusberg R.;
 RT Submitted (FEB-2001) to the EMBL/Genbank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =

```

CC      ubiquitin + a thiol.
CC      -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may
CC      be produced by alternative splicing.
CC      -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL: AF079564; AAC28392.1; -
DR      EMBL: BC002854; AAH02854.1; -
DR      EMBL: BC002955; AAH02955.1; -
DR      MEROPL: C19.013; -
DR      Genew: HGNC:12618; USP2.
DR      MIM: 604725; -
DR      InterPro: IPR001394; UCH-2.
DR      Pfam: PF00442; UCH-1; 1.
DR      Pfam: PF00443; UCH-2; 1.
DR      PROSITE: PS00972; UCH_2_1; 1.
DR      PROSITE: PS00973; UCH_2_2; 1.
DR      PROSITE: PS50235; UCH_2_3; 1.
DR      UBL conjugation pathway: Hydrolase; Thiol protease; Multigene family;
KW      Alternative splicing.
KM      ACT_SITE 24 24 BY SIMILARITY.
FT      ACT_SITE 297 297 BY SIMILARITY.
FT      ACT_SITE 305 305 BY SIMILARITY.
FT      VARSPLIC 1 252 MISSING (IN ISOFORM 2).
FT      VARSPLIC 253 258 PGRDGM -> MLNKKK (IN ISOFORM 2).
FT      CONFLICT 594 594 L -> H (IN REF. 1).
FT      CONFLICT 602 605 PSRM -> TSPI (IN REF. 1).
SQ      SEQUENCE 605 AA; 68071 MW; AAFPDA9344D21812 CRC64;

Query Match      54.7%; Score 35; DB 1; Length 605;
Best Local Similarity 70.0%; Pred. No. 54;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY      1 SPOGIAGGRN 10
DB      262 SAGLAGLRN 271

RESULT 22
ANDR_EULFC      STANDARD:      PRT;      884 AA.
AC      097776;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      Androgen receptor (Dihydrotestosterone receptor).
GN      AR OR NR3C4.
OS      Eulemur fulvus collaris (Collared brown lemur).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Strepsirrhini; Lemnidae; Eulemur.
OX      NCBI_TaxID=47178;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=98404153; Pubmed=9732460;
RA      Choong C.S., Kempainen J.A., Wilson E.M.;
RT      "Evolution of the primate androgen receptor: a structural basis for
RT      disease."
RL      J. Mol. Evol. 47:334-342(1998).
CC      -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC      THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC      PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC      -1- SUBCELLULAR LOCATION: Nucleat.
CC      -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC      A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC      -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC      NR3 SUBFAMILY.

```

```

CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL: U94178; AAC73049.1; -
DR      HSSP: P06536; 1DCD.
DR      TRASNSEC: T04655; -
DR      InterPro: IPR001103; Andrgn_receptor.
DR      InterPro: IPR000536; Hormone_rec_1lg.
DR      InterPro: IPR001628; znf_C4steroid.
DR      Pfam: PF00104; hormone_rec_1.
DR      Pfam: PF00105; znf_C4; 1.
DR      Pfam: PF02166; Androgen_recep; 1.
DR      PRINTS: PR00047; STROIDFINGER.
DR      ProDom: PD000035; znf_C4steroid; 1.
DR      SMART: SM00430; HOL1; 1.
DR      SMART: SM00399; znf_C4; 1.
DR      PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
DR      Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW      Zinc-finger; Steroid-binding.
KM      DNA_BIND 1 522 MODULATING (BY SIMILARITY).
FT      DNA_BIND 524 589 NUCLEAR RECEPTOR-TYPE.
FT      ZN_FING 524 544 C4-TYPE.
FT      ZN_FING 560 584 C4-TYPE.
FT      ZN_FING 560 584 C4-TYPE.
FT      DOMAIN 55 58 LIGAND-BINDING.
FT      DOMAIN 55 58 POLY-GLN.
FT      DOMAIN 64 70 POLY-GLN.
FT      DOMAIN 116 120 POLY-GLN.
FT      DOMAIN 174 178 POLY-GLN.
FT      DOMAIN 353 362 POLY-PRO.
FT      DOMAIN 379 383 POLY-ALA.
FT      DOMAIN 408 411 POLY-ALA.
FT      DOMAIN 430 435 POLY-GLY.
SQ      SEQUENCE 884 AA; 95610 MW; 18F570E352F4D2BD CRC64;

Query Match      54.7%; Score 35; DB 1; Length 884;
Best Local Similarity 85.7%; Pred. No. 80;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      2 POGIAGQ 8
DB      451 POGLAGQ 457

RESULT 23
ANDR_MACFA      STANDARD:      PRT;      895 AA.
AC      097952;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      Androgen receptor (Dihydrotestosterone receptor).
GN      AR OR NR3C4.
OS      Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC      Cercopithecinae; Macaca.
OX      NCBI_TaxID=9541;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=98404153; Pubmed=9732460;
RA      Choong C.S., Kempainen J.A., Wilson E.M.;
RT      "Evolution of the primate androgen receptor: a structural basis for
RT      disease."
RL      J. Mol. Evol. 47:334-342(1998).
CC      -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC      THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC      PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.

```

```

CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC NR3 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U94179; AAC73050.1; -.
CC HSSP: P06536; IGDC.
CC TRANSFAC: T04654; -.
CC InterPro: IPR001103; Andrgn_receptor.
CC InterPro: IPR000536; Hormone_rec_lig.
CC InterPro: IPR001628; Znfc4steroid.
CC Pfam: PF00104; hormone_rec.1.
CC Pfam: PF00105; zf-C4; 1.
CC Pfam: PF02166; Androgen_recep.1.
CC PRINTS: PR00047; STEROIDFINGER.
CC PRODOM: PD000035; Znfc4steroid.1.
CC SMART: SM00430; HOL1; 1.
CC SMART: SM00399; ZNF_C4; 1.
CC PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
CC Receptor: Transcription regulation; DNA-binding; Nuclear protein;
CC Zinc-finger; Steroid-binding.
CC -----
CC FT DNA_BIND 535 600 MODULATING (BY SIMILARITY).
CC FT ZN_FING 535 555 NUCLEAR RECEPTOR-TYPE.
CC FT ZN_FING 571 595 C4-TYPE.
CC FT DOMAIN 666 895 LIGAND-BINDING.
CC FT DOMAIN 55 62 POLY-GLN.
CC FT DOMAIN 68 74 POLY-GLN.
CC FT DOMAIN 178 182 POLY-GLN.
CC FT DOMAIN 357 366 POLY-PRO.
CC FT DOMAIN 381 387 POLY-ALA.
CC FT DOMAIN 434 448 POLY-GLY.
CC SEQUENCE 895 AA; 96494 MW; A3EB17916F43A097 CRC64;
SQ
Query Match 54.7%; Score 35; DB 1; Length 895;
Best Local Similarity 85.7%; Pred. No. 81;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 2 POGIAGO 8
Db 462 POGIAGO 468

```

```

CC -1- FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC NR3 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U94176; AAC73047.1; -.
CC HSSP: P06536; IGDC.
CC TRANSFAC: T04652; -.
CC InterPro: IPR001103; Andrgn_receptor.
CC InterPro: IPR000536; Hormone_rec_lig.
CC InterPro: IPR001628; Znfc4steroid.
CC Pfam: PF00104; hormone_rec.1.
CC Pfam: PF00105; zf-C4; 1.
CC Pfam: PF02166; Androgen_recep.1.
CC PRINTS: PR00047; STEROIDFINGER.
CC PRODOM: PD000035; Znfc4steroid.1.
CC SMART: SM00430; HOL1; 1.
CC SMART: SM00399; ZNF_C4; 1.
CC PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
CC Receptor: Transcription regulation; DNA-binding; Nuclear protein;
CC Zinc-finger; Steroid-binding.
CC -----
CC FT DNA_BIND 535 600 MODULATING (BY SIMILARITY).
CC FT ZN_FING 535 555 NUCLEAR RECEPTOR-TYPE.
CC FT ZN_FING 571 595 C4-TYPE.
CC FT DOMAIN 666 895 LIGAND-BINDING.
CC FT DOMAIN 55 63 POLY-GLN.
CC FT DOMAIN 69 74 POLY-GLN.
CC FT DOMAIN 178 182 POLY-GLN.
CC FT DOMAIN 357 366 POLY-PRO.
CC FT DOMAIN 381 387 POLY-ALA.
CC FT DOMAIN 434 448 POLY-GLY.
CC SEQUENCE 895 AA; 96478 MW; 9020C0DC67F11E5D CRC64;
SQ
Query Match 54.7%; Score 35; DB 1; Length 895;
Best Local Similarity 85.7%; Pred. No. 81;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 2 POGIAGO 8
Db 462 POGIAGO 468

```

```

RESULT 24
ID ANDR_PAPHA STANDARD; PRT; 895 AA.
AC 097960;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Androgen receptor (Dihydrotestosterone receptor).
GN AR OR NR3C4.
OS Papio hamadryas (Hamadryas baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Papio.
OX NCBI_TaxID=9557;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98404153; PubMed=9732460;
RA Choong C.S., Kempainen J.A., Wilson E.M.;
RT "Evolution of the primate androgen receptor: a structural basis for
RL J. Mol. Evol. 47:334-342(1998).

```

```

RESULT 25
ID ANDR_CANFA STANDARD; PRT; 907 AA.
AC 09790;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Androgen receptor (Dihydrotestosterone receptor).
GN AR OR NR3C4.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21618348; PubMed=11768233;
RA Lu B., Smock S.L., Castleberry T.A., Owen T.A.;
RT "Molecular cloning and functional characterization of the canine

```

```

RA Choong C.S., Kempainen J.A., Milson E.M.;
RT "Evolution of the primate androgen receptor: a structural basis for
   disease.";
RL J. Mol. Evol. 47:334-342(1998).
CC - FUNCTION: THE STEROID HORMONES AND THEIR RECEPTORS ARE INVOLVED IN
CC THE REGULATION OF EUKARYOTIC GENE EXPRESSION AND AFFECT CELLULAR
CC PROLIFERATION AND DIFFERENTIATION IN TARGET TISSUES.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC - SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC NRS SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U94177; AAC73048.1; -.
DR HSSP: P06536; IGDC.
DR TRANSFAC: T04653; -.
DR InterPro: IPR001103; Andrgn_receptor.
DR InterPro: IPR000536; Hormone_rec_1fg.
DR InterPro: IPR001628; znf_C4steroid.
DR Pfam: PF00104; hormone_rec.1.
DR Pfam: PF00105; zf-C4; 1.
DR Pfam: PF02166; Androgen_recep. 1.
DR PRINTS: PR00047; STEROIDFINGER.
DR ProDom: PD000035; znf_C4steroid; 1.
DR SMART: SM00430; HOL1; 1.
DR SMART: SM00399; znf_C4; 1.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Steroid-binding.
FT DOMAIN 1 549 MODULATING (BY SIMILARITY).
FT FT DNN_BIND 551 616 NUCLEAR RECEPTOR-TYPE.
FT FT ZN_FING 551 571 C4-TYPE.
FT FT ZN_FING 587 611 C4-TYPE.
FT FT DOMAIN 682 911 LIGAND-BINDING.
FT FT DOMAIN 57 78 POLY-GLN.
FT FT DOMAIN 84 88 POLY-GLN.
FT FT DOMAIN 192 196 POLY-GLN.
FT FT DOMAIN 371 380 POLY-PRO.
FT FT DOMAIN 395 401 POLY-ALA.
FT FT DOMAIN 448 464 POLY-GLY.
SQ SEQUENCE 911 AA; 98402 MW; 601B9BD4E697DAA4 CRC64;

Query Match 54.7%; Score 35; DB 1; Length 911;
Best Local Similarity 85.7%; Pred. No. 82;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PQGIAGQ 8
   111:111
Db 478 PQGIAGQ 484

RESULT 27
POL2_TRSVR
ID POL2_TRSVR STANDARD; PRT; 1882 AA.
AC P25247;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE RNA2 polypeptin (207 kDa protein) [Contains: Coat protein].
OS Tomato ringspot virus (Isolate raspberry) (Tomrsv).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Comoviridae;
OC Nepovirus.
OC NCBI_TaxID=12281;
OC [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE=91311402; PubMed=1856689;
 RA Rott M.E., Tremaine J.H., Roehon D.M.;
 RT "Nucleotide sequence of tomato ringspot virus RNA-2.";
 RL J. Gen. Virol. 72:1505-1514(1991).
 CC -1- SIMILARITY: IDENTICAL FOR THE FIRST 132 AA, AND 75.3% IDENTICAL
 CC FOR THE NEXT 145 AA TO THE RNA1 POLYPROTEIN.
 CC -1- SIMILARITY: TO THE RNA2 POLYPROTEIN OF OTHER NEPOVIRUSES.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-122 IS THE
 CC INITIATOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; D12477; BAA02043.1; -.
 CC DR PIR; J01093; GNVYR.
 CC DR InterPro; IPR005054; Nepo_coat.
 CC DR InterPro; IPR005305; Nepo_coat_C.
 CC DR InterPro; IPR005306; Nepo_coat_N.
 CC DR Pfam; PF03381; Nepo_coat_1.
 CC DR Pfam; PF03688; Nepo_coat_C; 1.
 CC DR Pfam; PF03689; Nepo_coat_N; 1.
 CC DR Polyprotein; Coat protein; Repeat.
 CC FT CHAIN 1321 1882 COAT PROTEIN (POTENTIAL).
 CC FT DOMAIN 554 698 2.5 X TANDEM REPEATS, PRO-RICH.
 CC FT REPEAT 554 606 1.
 CC FT REPEAT 607 659 2.
 CC FT REPEAT 660 698 3 (INCOMPLETE AND APPROXIMATE).
 CC SQ SEQUENCE 1882 AA; 206802 MM; 0F8958B63AEBDD9D CRC64;
 CC -----
 CC Query Match 54.7%; Score 35; DB 1; Length 1882;
 CC Best Local Similarity 54.5%; Pred. No. 1.7e+02;
 CC Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 CC -----
 CC QY 2 PGIAGQGNFN 12
 CC 1: 111: 11
 CC Db 1665 PRTAGEQGFN 1675
 CC -----
 CC RESULT 28
 CC EOTA_HUMAN STANDARD; PRT; 97 AA.
 CC ID EOTA_HUMAN
 CC AC P51671; P50877; Q92490; Q92491;
 CC DT 01-OCT-1996 (Rel. 34, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
 CC DE Eotaxin precursor (Small Inducible cytokine All) (CCL11) (Eosinophil
 CC DE chemotactic protein).
 CC GN SCYAL1.
 CC OS Homo sapiens (Human).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC OX NCBI_TaxID=9606;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=96181758; PubMed=8597956;
 CC RA Garcia-Zepeda E.A., Rothenberg M.E., Ombey T.R., Leder P.,
 CC Luster A.D.;
 CC RT "Human eotaxin is a specific chemoattractant for eosinophil cells and
 CC RT provides a new mechanism to explain tissue eosinophilia.";
 CC RL Nat. Med. 2:449-456(1996).
 CC [2]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=96189937; PubMed=8609214;
 CC RA Ponath P.D., Qin S., Ringle D.J., Clark-Lewis I., Wang J., Kassam N.,
 CC Smith H., Shi X., Gonzalo J.A., Newman W., Gutierrez-Ramos J.C.,
 CC Mackay C.R.;
 CC RT "Cloning of the human eosinophil chemoattractant, eotaxin. Expression,
 CC RT receptor binding, and functional properties suggest a mechanism for

RT the selective recruitment of eosinophils.";
 RL J. Clin. Invest. 97:604-612(1996).
 RL [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Small intestine;
 RX MEDLINE=96205964; PubMed=8631813;
 RA Kitaura M., Nakajima T., Imai T., Harada S., Combadiere C.,
 RA Tiffany H.L., Murphy P.M., Yoshie O.;
 RT "Molecular cloning of human eotaxin, an eosinophil-selective CC
 RT chemokine, and identification of a specific eosinophil eotaxin
 RT receptor, CC chemokine receptor 3.";
 RL J. Biol. Chem. 271:7725-7730(1996).
 RL [4]
 RP SEQUENCE FROM N.A., SEQUENCE OF 60-65 AND 75-88, AND VARIANTS.
 RC TISSUE=Fore skin;
 RX MEDLINE=96374440; PubMed=8780731;
 RA Bartels J., Schlueter C., Richter E., Noso N., Kulke R.,
 RA Christophers E., Schroeder J.M.;
 RT "Human dermal fibroblasts express eotaxin: molecular cloning, mRNA
 RT expression, and identification of eotaxin sequence variants.";
 RL Biochem. Biophys. Res. Commun. 225:1045-1051(1996).
 RL [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=97312708; PubMed=9169149;
 RA Garcia-Zepeda E.A., Rothenberg M.E., Wieremowicz S., Sarafi M.N.,
 RA Morton C.C., Luster A.D.;
 RT "Genomic organization, complete sequence, and chromosomal location of
 RT the gene for human eotaxin (SCYAL1), an eosinophil-specific CC
 RT chemokine.";
 RL Genomics 41:471-476(1997).
 RL [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=97445071; PubMed=9299399;
 RA Hein H., Schlueter C., Kulke R., Christophers E., Schroeder J.M.,
 RA Bartels J.;
 RT "Genomic organization, sequence, and transcriptional regulation of
 RT the human eotaxin gene.";
 RL Biochem. Biophys. Res. Commun. 237:537-542(1997).
 RL [7]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RA Strusberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RL [8]
 RP CARBOHYDRATE-LINKAGE SITE.
 RC TISSUE=Blood;
 RX MEDLINE=98237580; PubMed=9578468;
 RA Noso N., Bartels J., Mallet A.I., Mochizuki M., Christophers E.,
 RA Schroeder J.M.;
 RT "Delayed production of biologically active O-glycosylated forms of
 RT human eotaxin by tumor-necrosis-factor-alpha-stimulated dermal
 RT fibroblasts.";
 RL Eur. J. Biochem. 253:114-122(1998).
 RL [9]
 RP STRUCTURE BY NMR.
 RX MEDLINE=98380469; PubMed=9712872;
 RA Grump M.P., Rajaratnam K., Kim K.S., Clark-Lewis I., Sykes B.D.;
 RT "Solution structure of eotaxin, a chemokine that selectively recruits
 RT eosinophils in allergic inflammation.";
 RL J. Biol. Chem. 273:22471-22479(1998).
 CC -1- FUNCTION: IN RESPONSE TO THE PRESENCE OF ALLERGENS, THIS PROTEIN
 CC DIRECTLY PROMOTES THE ACCUMULATION OF EOSINOPHILS, A PROMINENT
 CC FEATURE OF ALLERGIC INFLAMMATORY REACTIONS. BINDS TO CCR3.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- INDUCTION: BY TNF-ALPHA, INTERLEUKIN-1 ALPHA AND INTERFERON GAMMA.
 CC -1- PTM: O-LINKED GLYCAN CONSISTS OF A GAL-GALNAc DISACCHARIDE WHICH
 CC IS MODIFIED WITH UP TO 2 STATIC ACID RESIDUES.
 CC -1- SIMILARITY: BELONGS TO THE INTERCINE BETA FAMILY (SMALL CYTOKINE
 CC C-C) (CHEMOKINE CC).
 CC -1- DATABASE: NAME=Ref Systems' cytokine source book: SCYAL1;
 CC WWW="http://www.indsystems.com/aasp/g_sitebuilder.asp?bodyid=196".

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U46573; AAA98957.1; -
DR EMBL: U34780; AAC50369.1; -
DR EMBL: D49372; BAA08370.1; -
DR EMBL: D69291; CAA93258.1; -
DR EMBL: Z75668; CAA99997.1; -
DR EMBL: Z75669; CAA99998.1; -
DR EMBL: U46572; AAC51297.1; -
DR EMBL: Z92709; CAB07027.1; -
DR EMBL: BC017850; AAA17850.1; -
DR PDB: 1EOT; 13-JUN-99.
DR PDB: 2EOT; 11-NOV-98.
DR Genew; HGNC:10610; SCYA11.
DR InterPro: IPR000827; CC_chemkine_sml.
DR InterPro: IPR001811; Chemokine_IL8.
DR Pfam; PF00048; IL8; 1.
DR SMART; SM00199; SCY; 1.
DR PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
DR Eosinophil; Cytokine; Chemotaxis; Glycoprotein; Signal;
KW Inflammatory response; Polymorphism; 3D-structure.
FT SIGNAL 1 23
FT CHAIN 24 97
FT DISULFID 32 57
FT DISULFID 33 73
FT CARBOHYD 94 94
FT VARIANT 7 7
FT VARIANT 23 23
FT VARIANT 51 51
FT VARIANT 79 79
FT SEQUENCE 97 AA; 10732 MW; B433C30FDAC71A7 CRC64;

Query Match 53.1%; Score 34; DB 1; Length 97;
Best Local Similarity 85.7%; Pred. No. 13;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SPOGIAG 7
DB 18 SPOGIAG 24

RESULT 29
STR KLEPN STANDARD; PRT; 266 AA.
AC P13082;
ID 01-JAN-1990 (rel. 13, Created)
DT 01-JAN-1990 (rel. 13, Last sequence update)
DT 01-OCT-1994 (rel. 30, Last annotation update)
DE Streptomycin 3'-kinase (EC 2.7.1.87) (Streptomycin 3'-
DE phosphotransferase) (SPH).
GN STR.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_Taxid=573;
RN [1]
RX MEDLINE=85215465; PubMed=3889831;
RA Mazodier P., Cossart P., Girard E., Gasser F.,
RT "Completion of the nucleotide sequence of the central region of Tns
RT confirms the presence of three resistance genes."

```

```

RL Nucleic Acids Res. 13:195-205(1985).
CC -1- FUNCTION: THE AMINOGLYCOSIDE PHOSPHOTRANSFERASES ACHIEVE
CC INACTIVATION OF THEIR ANTIBIOTIC SUBSTRATES BY PHOSPHORYLATION.
CC -1- CATALYTIC ACTIVITY: ATP + streptomycin -> ADP + streptomycin 3'-
CC phosphate.
CC -1- MISCELLANEOUS: THIS ENZYME IS ENCODED BY THE KANMYCIN AND
CC NEOMYCIN RESISTANCE TRANSPOSON TNS.
CC -1- SIMILARITY: TO OTHER AMINOGLYCOSIDE PHOSPHOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U00004; AAA73392.1; -
DR EMBL: X01702; CAA25854.1; -
DR PIR: B23034; B23034.
KW Antibiotic resistance; Transferase; Kinase; ATP-binding;
KW Transposable element.
FT ACT_SITE 154 154
FT ACT_SITE 154 154
FT SEQUENCE 266 AA; 29066 MW; 30A9DA5D0E1A826C CRC64;

Query Match 53.1%; Score 34; DB 1; Length 266;
Best Local Similarity 45.5%; Pred. No. 36;
Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 POGIAGORNFN 12
DB 173 PHGLGERFED 183

RESULT 30
YJ3C_SCHPO STANDARD; PRT; 278 AA.
AC 094404;
ID 15-JUN-2002 (rel. 41, Created)
DT 15-JUN-2002 (rel. 41, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Hypothetical protein C126.12 in chromosome III.
DE SPC126.12.
GN Schizosaccharomyces pombe (Fission yeast).
OS Schizosaccharomycetes pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_Taxid=4896;
RN [1]
RX MEDLINE=21846401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher G.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hildalgo J., Hodgson K.,
RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voicikert G., Aert R., Robben J., Gymnopoulos B.,
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moesti D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Wambuit R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leclaire V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Roeder M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

```

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schistosoma mansoni pombe."
RT Nature 415:871-880(2002)
CC -1- SIMILARITY: BELONGS TO THE UPF0135 (NIF3) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL034490; CAA22481.1; -
DR InterPro: IPR002678; DUF34.
DR Pfam: PFO1784; DUF34; 1.
DR TIGRFAMs: TIGR00486; DUF34; 1.
DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; 1.
DR Hypothetical protein.
KW SEQUENCE 278 AA; 29850 MW; A3314E79092D677C CRC64;
SQ
Query Match 53.1%; Score 34; DB 1; Length 278;
Best Local Similarity 87.5%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 3 QGAGGGRN 10
DB 122 QGAGGGRN 129

RESULT 31
YHWO_YEAST STANDARD; PRT; 283 AA.
ID YHWO_YEAST
AC P38855;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 32.0 kDa protein in REC104-SOL3, intergenic region.
GN YH160C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
RA Man M., Rikkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII."
RT Science 265:2077-2082(1994).
CC -1- SIMILARITY: SOME, TO YEAST PRT11.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U10397; AAB68992.1; -
DR PIR: S46770; S46770.
DR SGD: S0001203; PEX18.
KW Hypothetical protein.

SQ SEQUENCE 283 AA; 32037 MW; 574C644FD7ED7A71 CRC64;
Query Match 53.1%; Score 34; DB 1; Length 283;
Best Local Similarity 54.5%; Pred. No. 38;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 2 PGAGGGRN 12
DB 130 PGAGGGRN 140

RESULT 32
O85D_DROME STANDARD; PRT; 412 AA.
ID O85D_DROME
AC Q9VHQ2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative odorant receptor 85d.
GN O85D OR CG11742.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertone P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dong L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Milshina N.V., Modarres C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazkzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiter K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Sidenkios I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitek R., Tector C., Turner R., Venter R., Wang A., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RT Science 281:2185-2195(2000).
CC -1- FUNCTION: PROBABLE ROLE IN THE ODORANT RESPONSE, BEING AN ODORANT
CC RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO FAMILY DR-OR OF G-PROTEIN COUPLED
CC RECEPTORS.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AE003679; AAF54249.1;
 DR FlyBase: FBgn0037594; Or65d.
 DR InterPro: IPR004117; 7tm_6.
 DR Pfam: PF02949; 7tm_6; 1.
 KW Hypothetical protein; Transmembrane; G-protein coupled receptor;
 KW Olfaction; Multigene family.

FT DOMAIN 1 56 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 57 77 1 (POTENTIAL).
 FT DOMAIN 78 84 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 85 105 2 (POTENTIAL).
 FT DOMAIN 106 152 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 153 173 3 (POTENTIAL).
 FT DOMAIN 174 219 4 (POTENTIAL).
 FT TRANSMEM 220 240 4 (POTENTIAL).
 FT DOMAIN 241 282 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 283 303 5 (POTENTIAL).
 FT DOMAIN 304 314 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 315 335 6 (POTENTIAL).
 FT DOMAIN 336 382 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 383 403 7 (POTENTIAL).
 FT DOMAIN 404 412 CYTOPLASMIC (POTENTIAL).
 SO SEQUENCE 412 AA; 47987 MW; 0E4BE85F6C149EC CRC64;

Query Match
 Best Local Similarity 53.1%; Score 34; DB 1; Length 412;
 Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 POGIACORNFN 12
 DB 127 POGIACORNFN 137

RESULT 33
 UVRC_METH STANDARD: PRT; 579 AA.
 AC 026541;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Exonuclease ABC subunit C.
 GN UVRC OR MTH441.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
 OC Methanobacteriaceae; Methanothermobacter.
 OX NCBI_TaxID=187420;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Delta H;
 RX MEDLINE=98037514; PubMed=9371463;
 RA Smith D.R., Doucette-Stamm L.A., Delonghery C., Lee H.-M., Dubois J.,
 RA Aldege T., Bashirzadeh R., Blakey D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lum W., Pothier H., Qiu D.,
 RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT deltaH: functional analysis and comparative genomics.";
 RL J. Bacteriol. 179:7135-7155(1997).
 CC -1- FUNCTION: THE ABC EXCISION NUCLEASE IS A DNA REPAIR ENZYME THAT
 CC CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
 CC PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). ATTACHES TO THE
 CC UVA-UVB COMPLEX, DISPLACING THE UVA, AND THE DAMAGED DNA STRAND IS
 CC NICKED ON BOTH SIDES OF THE DAMAGED SITE (BY SIMILARITY).
 CC -1- SUBUNIT: CONSISTS OF THREE SUBUNITS; UVRB, UVRB AND UVRB.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE UVRC FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 UVR DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AE000828; AAB84947.1;
 DR HSPSP; P07025; 1E52.
 DR InterPro: IPR003583; HHH_1.
 DR InterPro: IPR000445; HHH.
 DR InterPro: IPR001943; UVRB/C.
 DR InterPro: IPR004791; UVR.C.
 DR InterPro: IPR001162; UVR.C.
 DR InterPro: IPR000305; UVR_C.N.
 DR Pfam: PF00633; HHH; 1.
 DR Pfam: PF01541; Excl_endo_N; 1.
 DR Pfam: PF02151; UVR; 1.
 DR ProDom: PD005870; UVR_C; 1.
 DR SMART: SM00465; GYVC; 1.
 DR SMART: SM00278; HhH1; 2.
 DR TIGRfam: TIGR00194; UVR.C; 1.
 DR PROSITE: PS0151; UVR; 1.
 DR SOS response; Exonuclease; DNA repair; Complete proteome.
 KW DOMAIN 193 228 UVR.
 FT SEQUENCE 579 AA; 66293 MW; 83D3DF78F9E3A68 CRC64;

Query Match
 Best Local Similarity 58.3%; Score 34; DB 1; Length 579;
 Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 SPOGIACORNFN 12
 DB 329 SPEGAGRRLLN 340

RESULT 34
 DHBK_LYCES STANDARD: PRT; 594 AA.
 AC 004059;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Putative 3,4-dihydroxy-2-butanone kinase (EC 2.7.1.-).
 GN DHBK.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot;
 OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
 OX NCBI_TaxID=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-cv. UC82B;
 RA Herz S., Eberhardt S., Bacher A.;
 RT "L. esculentum mRNA for 3,4-dihydroxy-2-butanone kinase.";
 RT Submitted (Apr-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE DIHYDROXYACETONE KINASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 UVR DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: Y12090; CAAT7805.1;
 DR InterPro: IPR004006; DAK1.

DR InterPro: IPR004007; Dakt2.
 DR Pfam: PF02733; Dakt1; 1.
 DR Pfam: PF02734; Dakt2; 1.
 KW Transferrase; Kinase.
 SQ SEQUENCE 594 AA; 61800 MW; BA19CBEA81BE24AB CRC64;

Query Match 53.1%; Score 34; DB 1; Length 594;
 Best Local Similarity 75.0%; Pred. No. 82;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 POGIAGOR 9
 Db 150 POGIAGOR 157

RESULT 35
 SNX9_HUMAN STANDARD; PRT; 595 AA.
 ID SNX9_HUMAN STANDARD; PRT; 595 AA.
 AC Q9Y5X1; Q90JH6; Q90P20; Q9BS17;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sorting nexin 9 (SNX9) and PX domain-containing protein 1 (SDP1 protein).
 GN SNX9 OR SH3PX1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NC NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21378165; PubMed=11485546;
 RA Teasdale R.D., Loch D., Houghton F., Karlsson L., Gleeson P.A.;
 RT "A large family of endosome-localized proteins related to sorting nexin 1";
 RL Biochem. J. 358:7-16(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20002705; PubMed=10531379;
 RA Howard L., Nelson K.K., Maciewicz R.A., Blobel C.P.;
 RT "Interaction of the metalloprotease disintegrins MDC9 and MDC15 with two SH3 domain-containing proteins, endophilin 1 and SH3PX1";
 RL J. Biol. Chem. 274:31693-31699(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Zhang J.S., Smith D.I.;
 RT "Identification of differentially expressed genes in matched prostate cancer and normal epithelial cell lines";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TTSOE-Skin;
 RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE OF 101-549 FROM N.A.
 RA Almeida J.;
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE OF 201-595 FROM N.A.
 RA Ramanathan G., Subramaniam V.N., Hong W.;
 RL "Human SDP1";
 RT Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: May be involved in several stages of intracellular trafficking.
 CC -1- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 PX DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: AF121859; AAD27832.1;
 DR EMBL: AF131214; AAF04473.1;
 DR EMBL: AF172847; AAL54871.1;
 DR EMBL: BC005022; AAL05022.1;
 DR EMBL: AL035634; CAB46196.1;
 DR EMBL: AF076957; AAD43001.1;
 DR Genew: HGNC:14973; SNX9.
 DR MTM: 605952; SNX9.
 DR InterPro: IPR001683; PX.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00018; SH3; 1.
 DR Pfam: PF00787; PX; 1.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR PRODOM: PD000066; SH3; 1.
 DR SMART: SM00312; PX; 1.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS50195; PX; 1.
 DR PROSITE: PS50002; SH3; 1.
 KW Transport; Protein transport; SH3 domain.
 FT DOMAIN 1 62 SH3.
 FT DOMAIN 250 361 PX.
 FT CONFLICT 89 89 Q -> H (IN REF. 4).
 SQ SEQUENCE 595 AA; 66591 MW; 963892AC1A5A9227 CRC64;

Query Match 53.1%; Score 34; DB 1; Length 595;
 Best Local Similarity 54.5%; Pred. No. 83;
 Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 POGIAGORNFN 12
 Db 129 POGIAGORNFN 139

RESULT 36
 LEU2_RHIRA STANDARD; PRT; 644 AA.
 ID LEU2_RHIRA STANDARD; PRT; 644 AA.
 AC P17279;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE 3-Isopropylmalate dehydratase (EC 4.2.1.33) (Isopropylmalate isomerase) (Alpha-IPM isomerase) (IPMI).
 GN LEU2.
 OS Rhizomucor racemosus (Mucor circinelloides f. lusitanicus).
 OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Mucoraceae;
 CC Mucor.
 CC NCB1_TaxID=4841;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 1216B;
 RX MEDLINE=90128278; PubMed=2693214;
 RA Isabel M., Roncero G., Jensen L.P., Stroeman P., van Heeswijk R.;
 RT "Characterization of a leuA gene and an ARS element from Mucor circinelloides";
 RL Gene 84:335-343(1989).
 CC -1- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate and 3-isopropylmalate, via the formation of 2-isopropylmalate.
 CC -1- CATALYTIC ACTIVITY: 3-isopropylmalate = 2-isopropylmalate + H(2)O.
 CC -1- CATALYTIC ACTIVITY: 2-isopropylmalate + H(2)O = 2-isopropylmalate.
 CC -1- PATHWAY: Leucine biosynthesis; second step.
 CC -1- SUBUNIT: Monomer.
 CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.lsb-sib.ch/announce/> or send an email to license@lsb-sib.ch).

CC EMBL: M31166; AAA33422.1; -
 DR PIR: J00160; J00160.
 DR PIR: S26865; S26865.
 DR InterPro: IPR000573; Aconitase_C.
 DR InterPro: IPR001030; Aconitase_N.
 DR InterPro: IPR004430; Leuc.
 DR Pfam: PF00330; Aconitase_1.
 DR Pfam: PF00694; Aconitase_C; 1.
 DR PRINTS: PR00415; ACONITASE.
 DR PRODOM: PD000511; Aconitase_N; 1.
 DR TIGRfams: TIGR00170; Leuc; 1.
 DR PROSITE: PS00450; ACONITASE_1; 1.
 DR PROSITE: PS01244; ACONITASE_2; 1.
 KW Leucine biosynthesis; Lyase; Iron-sulfur; 4Fe-4S; (BY SIMILARITY).
 FT METAL 400 400 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 460 460 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 463 463 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 SQ SEQUENCE 644 AA; 69939 MW; 087FC2C518BD497E CRC64;

Query Match 53.1%; Score 34; DB 1; Length 644;
 Best Local Similarity 58.3%; Pred. No. 90;
 Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 SPOCIAGORNFN 12
 DB 84 SPOAFEGLRNAN 95

RESULT 37
 ID TFE2_HUMAN STANDARD; PRT: 654 AA.
 AC P15883; O9UP19; Q14635; Q14636; Q14208;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last annotation update)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Transcription factor E2-alpha (Immunoglobulin enhancer binding factor E12/447) (Transcription factor-3) (TCF-3) (Immunoglobulin de transcription factor-1) (Transcription factor TTF-1) (Kappa-E2-binding factor).
 GN TCF3 OR E2A OR TTF1.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN RP SEQUENCE FROM N.A. (ISOFORM E12).
 RX MEDLINE-90150282; PubMed-19679983;
 RA Kamp M.P., Murte C., Sun X.-H., Baltimore D.;
 RT "A new homeobox gene contributes the DNA binding domain of the t(1;19) translocation protein in pre-B ALL.";
 RL Cell 60:547-555(1990).
 RN RP SEQUENCE FROM N.A. (ISOFORM E12).
 RX MEDLINE-90150281; PubMed-19679982;
 RA Nourse J., Mellentin J.D., Gallili N., Wilkinson J., Standridge E., Smith S.D., Cleary M.L.;
 RT "Chromosomal translocation t(1;19) results in synthesis of a homeobox fusion mRNA that codes for a potential chimeric transcription factor.";
 RL Cell 60:535-545(1990).
 RN RP SEQUENCE FROM N.A. (ISOFORM E12).
 RA Lamerding J.E., McCreedy P.M., Skowronski E., Viswanathan V., Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S., Phan H., Velasco N., Do L., Regala W., Terry A., Gaines J., Dangnan L., Ertel A., Christensen W., Georgescu A., Avila J., Liu S., Altix C., Andeise T., Frankheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B., Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,

RA Olsen A.S., Carrano A.V.;
 RT "Sequence analysis of a 3.5 Mb contig in human 19p13.3 containing a serine protease gene cluster.";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RP [4]
 RP SEQUENCE OF 214-654 FROM N.A. (ISOFORMS E12 AND E47), AND BHLH DOMAIN.
 RC TISSUE=Lymphoma;
 RX MEDLINE-89168418; PubMed-2493990;
 RA Murte C., McCaw P.S., Baltimore D.;
 RT "A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, Myod, and myc proteins.";
 RL Cell 56:777-783(1989).
 RN [5]
 RP SEQUENCE OF 69-654 FROM N.A. (ISOFORM E47).
 RX MEDLINE-90175015; PubMed-2308859;
 RA Henthorn P., McCarrick-Walmsley R., Kadesch T.;
 RT "Sequence of the cDNA encoding TTF-1, a positive-acting transcription factor.";
 RL Nucleic Acids Res. 18:677-677(1990).
 RN [6]
 RP SEQUENCE OF 511-654 FROM N.A. (ISOFORM E47).
 RX MEDLINE-92297964; PubMed-1818757;
 RA Zhang Y., Bina M.;
 RT "Sequence of a Hela cDNA provides the DNA binding domain and carboxy terminus of HE47: a human helix-loop-helix protein related to the enhancer binding factor E47.";
 RL DNA Seq. 2:197-202(1992).
 RN [7]
 RP DISCUSSION OF SEQUENCE.
 RX MEDLINE-90140708; PubMed-2105528;
 RA Henthorn P., Kiledjian M., Kadesch T.;
 RT "Two distinct transcription factors that bind the immunoglobulin enhancer microE5/Kappa 2 motif.";
 RL Science 247:467-470(1990).
 RN [8]
 RP MUTAGENESIS.
 RX MEDLINE-90280447; PubMed-2112746;
 RA Voronova A., Baltimore D.;
 RT "Mutations that disrupt DNA binding and dimer formation in the E47 helix-loop-helix protein map to distinct domains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:4722-4726(1990).
 CC -1- FUNCTION: HETERODIMERS BETWEEN TCF3 AND TISSUE-SPECIFIC BASIC TISSUE-SPECIFIC CELL FATE DURING EMBRYOGENESIS. LIKE MUSCLE OR EARLY B-CELL DIFFERENTIATION. DIMERS BIND DNA ON E-BOX MOTIFS: 5'-CANNCTG-3'. BINDS TO THE KAPPA-E2 SITE IN THE KAPPA IMMUGLOBULIN GENE ENHANCER.
 CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER BHLH PROTEIN. FORMS A HETERODIMER WITH ASH1.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: E47/PAN-1 AND E12/PAN-2 (SHOWN HERE). ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- PTM: PHOSPHORYLATED FOLLOWING NGF STIMULATION (BY SIMILARITY).
 CC -1- DISEASE: A FORM OF PRE-B-CELL ACUTE LYMPHOBLASTIC LEUKEMIA (B-ALL) IS CHARACTERIZED BY A CHROMOSOMAL TRANSLOCATION T(1;19)(O2;P13.3) WHICH INVOLVES PBX1 AND TCF3.
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF TRANSCRIPTION FACTORS.
 CC -1- DATABASE: NAME=Atlas genet. Cytogenet. Oncol. Haematol.;
 CC WWW="http://www.infobiogen.fr/services/chromocancer/genes/E2A.html".
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation-CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.lsb-sib.ch/announce/> or send an email to license@lsb-sib.ch).

CC EMBL: M31523; AAA61146.1; -
 DR EMBL: M31522; AAA36764.1; ALT-SDO.
 DR EMBL: M31222; AAA52331.1; ALT-INTT.

DR EMBL: AC006274; AAC99797.1; -
 DR EMBL: AC005321; AAC27373.1; -
 DR EMBL: M24404; AAA56829.1; -
 DR EMBL: M24405; AAA56830.1; -
 DR EMBL: X52078; CAA36297.1; -
 DR EMBL: M65214; AAC1693.1; -
 DR PIR: A31492; A31492.
 DR PIR: A34734; A34734.
 DR PIR: S10099; S10099.
 DR TRANSFAC: T00204; -
 DR Genew: HGNC:11633; TCF3.
 DR MIM: 147141; -
 DR InterPro: IPR001092; HLH_basic.
 DR Pfam: PF00010; HLH_1.
 DR SMART: SM00353; HLH_1.
 DR PROSITE: PS00038; HLH_1; 1.
 DR PROSITE: PS50888; HLH_2; 1.
 KW Transcription regulation; DNA-binding; Nuclear protein;
 KW Proto-oncogene; Chromosomal translocation; Alternative splicing;
 KW Phosphorylation.
 FT DOMAIN 176 176
 FT DOMAIN 389 425
 FT DNA_BIND 547 561
 FT DOMAIN 562 605
 FT SITE 483 484
 FT VARSPLIC 530 601
 FT
 FT MUTAGEN 550 551
 FT MUTAGEN 551 551
 FT MUTAGEN 561 561
 FT MUTAGEN 561 563
 FT MUTAGEN 563 588
 FT MUTAGEN 588 588
 FT MUTAGEN 591 592
 FT MUTAGEN 595 595
 FT
 FT CONFLICT 69 99
 FT CONFLICT 214 216
 FT CONFLICT 390 390
 FT CONFLICT 552 552
 FT CONFLICT 560 560
 FT CONFLICT 570 570
 FT CONFLICT 578 578
 FT CONFLICT 581 585
 FT CONFLICT 593 593
 FT CONFLICT 597 597
 FT CONFLICT 601 601
 FT SEQUENCE 654 AA; 67600 MM; 52F5E3DE1890AET3 CRC64;
 SO
 Query Match 53.18; Score 34; DB 1; Length 654;
 Best Local Similarity 85.7%; Pred. No. 91;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OC NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Foreskin;
 RA MEDLINE=98370844; PubMed=9707347;
 RA Pinhajama T., Vuorio M.M., Annunen S., Perelae M., Prockop D.J.,
 RA Ala-Kokko L.,
 RT "Human COL9A1 and COL9A2 genes. Two genes of 90 and 15 kb code for
 RT similar polypeptides of the same collagen molecule."
 RL Matrix Biol. 17:237-241(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Donnelly S.,
 RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 50-668 FROM N.A.
 RC TISSUE=Cartilage;
 RA MEDLINE=93202262; PubMed=8454052;
 RA Perelae M., Hanninen M., Hastbacka J., Elima K., Vuorio E.,
 RT "Molecular cloning of the human alpha 2(I) collagen cDNA and
 RT assignment of the human COL9A2 gene to chromosome 1."
 RL FEBS Lett. 319:177-180(1993).
 RN [4]
 RP VARIANT 100 TRP-326, AND VARIANT ARG-326.
 RA PubMed=10411504;
 RA Annunen S., Paasilta P., Lohiiva J., Perelae M., Pinhajama T.,
 RA Karppinen J., Tervonen O., Kroeger H., Laeide S., Vanharanta H.,
 RA Ryhanen L., Goering H.H., Oht J., Prockop D.J., Ala-Kokko L.,
 RT "An allele of COL9A2 associated with intervertebral disc disease."
 RL Science 285:409-412(1999).
 CC
 CC -1- FUNCTION: Structural component of hyaline cartilage and vitreous
 CC of the eye.
 CC
 CC -1- SUBUNIT: Heterotrimer of an alpha 1(I), an alpha 2(I) and an alpha
 CC 3(I) chain.
 CC
 CC -1- PTM: Covalently linked to the telopeptides of type II collagen by
 CC lysine-derived cross-links.
 CC
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC
 CC -1- DISEASE: Defects in COL9A2 are a cause of multiple epiphyseal
 CC dysplasia type 2 (EDM2). EDM2 is characterized by flattened,
 CC irregular epiphyses in most joints, particularly the knees.
 CC
 CC -1- DISEASE: Defects in COL9A2 are a cause of intervertebral disc
 CC disease (IDD).
 CC
 CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
 CC INTERRUPTED HELICES (FACIT) FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AF019406; AAC35512.1; -
 CC EMBL: AL050341; CAB81611.1; -
 DR EMBL: M95610; AAA80977.1; -
 DR Genew: HGNC:2218; COL9A2.
 DR MIM: 120260; -
 DR MIM: 600204; -
 DR MIM: 603932; -
 DR InterPro: IPR000087; Collagen.
 DR Pfam: PF01391; Collagen; 9.
 DR ProDom: PD000007; Collagen; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Cartilage; Collagen; Signal; Disease mutation; Polymorphism.
 FT SIGNAL 1 23
 FT CHAIN 24 689
 FT DOMAIN 27 519
 FT DOMAIN 520 549
 FT NONHELICAL REGION 3 (NC3).
 FT NONHELICAL REGION 3 (NC3).
 FT

```

FT DOMAIN 550 632 TRIPLE-HELICAL REGION 2 (COL2).
FT DOMAIN 633 634 NONHELICAL REGION 2 (NC2).
FT DOMAIN 635 664 TRIPLE-HELICAL REGION 3 (COL1).
FT DOMAIN 665 689 NONHELICAL REGION 1 (NC1).
FT VARIANT 326 326 Q -> R.
FT VARIANT 326 326 /FTID-VAR-012659.
FT VARIANT 326 326 O -> W (IN ID); REQUIRES 2 NUCLEOTIDE
FT VARIANT 326 326 SUBSTITUTIONS.
FT VARIANT 326 326 /FTID-VAR-012658.
SO SEQUENCE 689 AA; 65131 MW; EB6106E02FEFA62 CRC64;

Query Match
Best Local Similarity 53.1%; Score 34; DB 1; Length 689;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 POGIACOR 9
DB 380 POGIMGOK 387

RESULT 39
YDZB_SCHPO STANDARD; PRT; 734 AA.
ID YDZB_SCHPO
AC 013718;
AD 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C14C4.11 in chromosome I.
GN SPAC14C4.11.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OC NCBI_TaxId=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grynpert B.,
RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Meller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Leinrich H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Wambolt R., Purrelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleuvre V., Mottier S.,
RA Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garçon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Useery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: TO YEAST YJL012C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

CC or send an email to license@isb-sib.ch.
CC
CC EMBL: Z98596; CAB11204.1; -.
CC InterPro: IPR003807; DUF202.
CC InterPro: IPR004331; SPX.
CC Pfam: PF02656; DUF202; 1.
CC Pfam: PF03105; SPX; 1.
CC Hypothetical protein; Transmembrane.
FT TRANSMEM 470 490 POTENTIAL.
FT TRANSMEM 625 645 POTENTIAL.
FT TRANSMEM 651 671 POTENTIAL.
FT TRANSMEM 694 714 POTENTIAL.
SO SEQUENCE 734 AA; 85445 MW; 2373306038C7F4EE CRC64;

Query Match
Best Local Similarity 53.1%; Score 34; DB 1; Length 734;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 POGIACOR 10
DB 505 PEGSAGSRN 513

RESULT 40
LEU2_USTMA
ID LEU2_USTMA STANDARD; PRT; 773 AA.
AC P49601.
AD 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 3-Isopropylmalate dehydratase (EC 4.2.1.33) (isopropylmalate
DE isomerase) (alpha-IPM isomerase) (IPMI).
GN LEU1.
OS Ustilago maydis (Smut fungus).
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OC NCBI_TaxId=5270;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=94171070; PubMed=8125330;
RX Rubin B.P., Li D., Holloman W.K.;
RT "The LEU1 gene of Ustilago maydis."
RL Gene 140:131-135(1994).
CC -1- FUNCTION: Catalyzes the isomerization between 2-isopropylmalate
CC and 3-isopropylmalate, via the formation of 2-isopropylmalate.
CC -1- CATALYTIC ACTIVITY: 3-isopropylmalate = 2-isopropylmalate +
CC H(2)O.
CC -1- CATALYTIC ACTIVITY: 2-isopropylmalate + H(2)O = 2-
CC isopropylmalate.
CC -1- PATHWAY: Leucine biosynthesis; second step.
CC -1- SUBUNIT: Monomer (by similarity).
CC -1- SIMILARITY: BELONGS TO THE ACONITASE/IPM ISOMERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch.
CC
CC EMBL: L20832; AAA34226.1; -.
CC InterPro: IPR000573; Aconitase_C.
CC InterPro: IPR001030; Aconitase_N.
CC InterPro: IPR004430; LeuC.
CC InterPro: IPR004431; LeuD.
CC Pfam: PF00330; aconitase; 1.
CC Pfam: PF00694; Aconitase_C; 1.
CC PRINTS: PR00415; ACONITASE.
CC PRODOM: PD000511; Aconitase_N; 1.
CC TIGRFAMs: TIGR00170; leuC; 1.
CC TIGRFAMs: TIGR00171; leuD; 1.
CC PROSITE: PS00450; ACONITASE_1; 1.

```

DR PROSITE: PS001244; ACONITASE_2; 1.
 KM Leucine biosynthesis: Lyase; Iron-sulfur; 4Fe-4S.
 FT METAL 355 355 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 415 415 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 FT METAL 418 418 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
 SQ SEQUENCE 773 AA; 83224 MW; A936D8D190DDC1C CRC64;
 Query Match 53.1%; Score 34; DB 1; Length 773;
 Best Local Similarity 58.3%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 SPOGIAGORNF 12
 Db 38 SPOAGFGRNAN 49
 RESULT 41
 ID SYL_CHLMT STANDARD; PRT; 819 AA.
 AC 09PK14;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
 GN LEUS OR TC0481.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MoPn / N199;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwynn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 pneumoniae A39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE001294; AAC67801.1;
 DR TIGR: TC0481;
 DR InterPro: IPR002302; Leu-TRNASyntla.
 DR InterPro: IPR002300; tRNA-synt_1a.
 DR InterPro: IPR001412; tRNA-synt_1.
 DR Pfam: PF00133; tRNA-synt_1; 1.
 DR PRINTS: PRO0985; TRNASYNTHLEU.
 DR TIGRFAMS: TIGR00396; leus_dact. 1.
 DR PROSITE: PS00178; AA-TRNA_LIGASE_I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT SITE 40 "HIGH" REGION.
 FT SITE 600 "KMSKS" REGION.
 FT BINDING 603 ATP (BY SIMILARITY).
 SQ SEQUENCE 819 AA; 93014 MW; 024D7E1BD2E500D4 CRC64;
 Query Match 53.1%; Score 34; DB 1; Length 819;
 Best Local Similarity 63.6%; Pred. No. 1.1e+02;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 SPOGIAGORNF 11
 Db 640 SNOGVAGCRRF 650
 RESULT 42
 ID SYL_CHLTR STANDARD; PRT; 819 AA.
 AC 084211;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
 GN LEUS OR CT209.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UW-3/Cx;
 RX MEDLINE=99000809; PubMed=9784136;
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatunov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis.";
 RL Science 282:754-759(1998).
 CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) = AMP +
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE001294; AAC67801.1;
 DR InterPro: IPR002302; Leu-TRNASyntla.
 DR InterPro: IPR002300; tRNA-synt_1a.
 DR InterPro: IPR001412; tRNA-synt_1.
 DR Pfam: PF00133; tRNA-synt_1; 1.
 DR PRINTS: PRO0985; TRNASYNTHLEU.
 DR TIGRFAMS: TIGR00396; leus_dact. 1.
 DR PROSITE: PS00178; AA-TRNA_LIGASE_I; 1.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 FT SITE 40 "HIGH" REGION.
 FT SITE 600 "KMSKS" REGION.
 FT BINDING 603 ATP (BY SIMILARITY).
 SQ SEQUENCE 819 AA; 92903 MW; 82412A659AFAEFCF CRC64;
 Query Match 53.1%; Score 34; DB 1; Length 819;
 Best Local Similarity 63.6%; Pred. No. 1.1e+02;
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 SPOGIAGORNF 11
 Db 640 SNOGVAGCRRF 650
 RESULT 43
 ID STCL_EMENI STANDARD; PRT; 1559 AA.
 AC 000681;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative sterigmacyclin biosynthesis fatty acid synthase alpha
 DE subunit.

RC SEQUENCE FROM N.A.
 RP STRAIN-H37rv:
 RA MEDLINE-98295987; PubMed-9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekela F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Dehail K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagers K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sutcliffe J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN-CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson S.J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ernolova M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Z77163; CAB00992.1; -
 DR EMBL: AE007078; AAK46649.1; -
 DR TIGR: MT2363; -
 DR TubercuList; RV2306c; -
 KW Hypothetical protein; signal; Complete proteome.
 FT SIGNAL 1 16 POTENTIAL.
 FT CHAIN 17 208 HYPOTHELTICAL PROTEIN RV2306c.
 FT FT
 SQ SEQUENCE 208 AA; 22521 MW; DE913A5D2C34DA2F CXC64;

 Query Match 51.6%; Score 33; DB 1; Length 208;
 Best Local Similarity 62.5%; Pred. No. 44;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 QY 2 POGINGOR 9
 Db 89 PEGVNGOR 96

 RESULT 45
 RS5_MYCLE
 ID RS5_MYCLE STANDARD; PRT; 217 AA.
 AC Q33000;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 16-JUL-1998 (Rel. 36, Last sequence update)
 DE 30S ribosomal protein S5.
 DE RESE OR ML1842 OR MCB82492.21.
 OS Mycobacterium lepreae.
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteriidae;
 OC Actinomycetales; Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI-TaxId=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TN;
 RA MEDLINE-2118732; PubMed-11234002;
 RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Nungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,

RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus";
 RL Nature 409:1007-1011(2001).
 CC -1- FUNCTION: PROTEIN SS IS IMPORTANT IN THE ASSEMBLY AND FUNCTION OF
 CC THE 30S RIBOSOMAL SUBUNIT (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE S5P FAMILY OF RIBOSOMAL PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: 298756; CAB11453.1; -
 CC EMBL: AL583923; CAC30796.1; -
 CC HSSP: P02357; 1PKP.
 CC Leptoma; ML1842; -
 CC DR InterPro: IPR000851; Ribosomal_S5.
 CC DR InterPro: IPR005324; Ribosomal_S5_C.
 CC DR Pfam: PF00333; Ribosomal_S5_1.
 CC DR Pfam: PF03719; Ribosomal_S5_C_1.
 CC DR TIGRfam: TIGR01021; rpsE_bact; 1.
 CC DR PROSITE: PS00385; RIBOSOMAL_S5; 1.
 CC KW Ribosomal protein; Complete proteome.
 CC SQ SEQUENCE 217 AA; 22614 MW; CE1D09563DC700B4 CRC64;
 CC
 CC Query Match 51.6%; Score 33; DB 1; Length 217;
 CC Best Local Similarity 55.6%; Pred. No. 46;
 CC Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 2 PQGIAGQRN 10
 CC I I I I I I I
 CC Db 197 PQAVAGERS 205
 CC
 CC RESULT 46
 CC HEM4_SYNP7 STANDARD; PRT; 264 AA.
 CC ID HEM4_SYNP7
 CC AC P42452;
 CC DT 01-NOV-1995 (Rel. 32, Created)
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
 CC DE Uroporphyrinogen-III synthase (EC 4.2.1.75) (UROS) (Uroporphyrinogen-
 CC III cosynthetase) (Hydroxymethylbilane hydrolyase [cyclizing]).
 CC GN HEMD.
 CC OS Synecococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
 CC OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
 CC OX NCBI_TaxID=1140;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RA MEDLINE-94169298; PubMed-8123787;
 CC RA Jones M.C., Jenkins J.M., Smith A.G., Howe C.J.;
 CC RT "Cloning and characterisation of genes for tetrapyrrole biosynthesis
 CC from the cyanobacterium Anacystis nidulans R2";
 CC RL Plant Mol. Biol. 24:435-448(1994).
 CC -1- CATALYTIC ACTIVITY: Hydroxymethylbilane - uroporphyrinogen-III +
 CC H(2O).
 CC -1- PATHWAY: Porphyrin biosynthesis; fourth step.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC

DR EMBL: X70966; CAA50303.1; -
 DR InterPro: IPR003754; HEMA.
 DR Pfam: PF02602; HEMA; 1.
 DR Porphyrin biosynthesis; Lyase.
 DR ACT_SITE 146 146 POTENTIAL.
 DR FT ACT_SITE 146 146
 DR SQ SEQUENCE 264 AA; 28196 MW; 0BAB84E9F8E80FA CRC64;
 CC
 CC Query Match 51.6%; Score 33; DB 1; Length 264;
 CC Best Local Similarity 75.0%; Pred. No. 56;
 CC Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 2 PQGIAGQR 9
 CC I I I I I I I
 CC Db 128 PQVAGQR 135
 CC
 CC RESULT 47
 CC CRKL_HUMAN STANDARD; PRT; 303 AA.
 CC ID CRKL_HUMAN
 CC AC P46109;
 CC DT 01-NOV-1995 (Rel. 32, Created)
 CC DT 01-NOV-1995 (Rel. 32, Last sequence update)
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
 CC DE Crk-like protein.
 CC GN CRKL.
 CC OS Homo sapiens (Human).
 CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC OX NCBI_TaxID=9606;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC TISSUE=Spleen;
 CC RX MEDLINE-93368949; PubMed-8361759;
 CC RA ten Hoeve J., Morris C., Heisterkamp N., Groffen J.;
 CC RT "Isolation and chromosomal localization of CRKL, a human crk-like
 CC gene";
 CC RL Oncogene 8:2469-2474(1993).
 CC -1- FUNCTION: MAY MEDIATE THE TRANSDUCTION OF INTRACELLULAR SIGNALS.
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X59656; CAA42199.1; -
 CC HSSP: Q64010; ICRA.
 CC DR Genew: HGNC:2363; CRKL.
 CC DR MIM: 602007; -
 CC DR InterPro: IPR000980; SH2.
 CC DR InterPro: IPR001452; SH3.
 CC DR Pfam: PF00017; SH2; 1.
 CC DR Pfam: PF00018; SH3; 1.
 CC DR PRINTS: PR00401; SH2DOMAIN.
 CC DR PRINTS: PR00452; SH3DOMAIN.
 CC DR ProDom: PD000066; SH3; 1.
 CC DR ProDom: PD000093; SH2; 1.
 CC DR SMART: SM00252; SH2; 1.
 CC DR SMART: SM00326; SH3; 2.
 CC DR PROSITE: PS0001; SH2; 1.
 CC DR PROSITE: PS0002; SH3; 1.
 CC DR SH2 domain; SH3 domain; Repeat.
 CC FT DOMAIN 14 102 SH2.
 CC FT DOMAIN 123 183 SH3 1.
 CC FT DOMAIN 235 296 SH3 2.
 CC SQ SEQUENCE 303 AA; 33777 MW; 294CF1EE2CD44B81 CRC64;
 CC
 CC Query Match 51.6%; Score 33; DB 1; Length 303;
 CC Best Local Similarity 58.3%; Pred. No. 64;
 CC

Matches 7: Conservative 0: Mismatches 5: Indels 0: Gaps 0:
 OY 1 SPOGIAGORNFN 12
 DB 185 SPHGKGNRNSN 196

RESULT 48
 CRL_MOUSE STANDARD: PRT: 303 AA.
 ID CRL_MOUSE
 AC P47941:
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Crl-like protein.
 GN CRL OR CRL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;

SEQUENCE FROM N.A.
 RA STRAIN-C57BL X CBA; TISSUE-Placenta;
 RX MEDLINE=96038874; PubMed=7478571;
 RA de Jong R.L., Haataja L., Voncken J.W., Heisterkamp N., Groffen J.;
 RT "Tyrosine phosphorylation of murine Crl.";
 RL Oncogene 11:1459-1474(1995).
 CC
 CC -1- FUNCTION: MAY MEDIATE THE TRANSDUCTION OF INTRACELLULAR SIGNALS.
 CC -1- P-TM: PHOSPHORYLATED ON TYROSINE. PHOSPHORYLATION IS PROMINENT
 CC DURING EARLY DEVELOPMENT, BUT DECREASES AT LATER EMBRYONIC STAGES
 CC AND IN NEBORN MICE.
 CC
 CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X90648; CAA62220.1;
 DR HSSP: Q64010; ICKA.
 DR MGSI:104686; Crlol.
 DR InterPro: IPR000980; SH2.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00017; SH2; 1.
 DR Pfam: PF00018; SH3; 1.
 DR PRINTS: PR00401; SH2DOMAIN.
 DR PRINTS: PR00452; SH3DOMAIN.
 DR PRODOM: PD000066; SH3; 1.
 DR PRODOM: PD000093; SH2; 1.
 DR SMART: SM00252; SH2; 1.
 DR SMART: SM00326; SH3; 2.
 DR PROSITE: PS50001; SH2; 1.
 DR PROSITE: PS50002; SH3; 1.
 KW SH2 domain; SH3 domain; Repeat; Phosphorylation.
 FT DOMAIN 14 102 SH2.
 FT DOMAIN 123 183 SH3 1.
 FT DOMAIN 225 296 SH3 2.
 SQ SEQUENCE 303 AA: 33817 MW: B8126DDDF2053573 CRC64;

Query Match 51.6%; Score 33; DB 1; Length 303;
 Best Local Similarity 58.3%; Pred. No. 64;
 Matches 7: Conservative 0: Mismatches 5: Indels 0: Gaps 0:
 OY 1 SPOGIAGORNFN 12
 DB 185 SPHGKGNRNSN 196

RESULT 49

GSHB_ANASP STANDARD: PRT: 324 AA.
 ID GSHB_ANASP
 AC P45480: Q43879;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Glutathione synthetase (EC 6.3.2.3) (Glutathione synthase) (GSH
 DE synthetase) (GSH-S) (GSHase).
 GN GSHB OR GSH-II OR AL3859.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103650;

SEQUENCE FROM N.A.
 RA MEDLINE=96001250; PubMed=7557485;
 RA Doherty H.M., Adams D.G.;
 RT "Cloning and sequence of ftsZ and flanking regions from the
 RL Gene 163:93-96(1995).
 CC
 CC [2]
 CC SEQUENCE FROM N.A.
 RA MEDLINE=21595285; PubMed=11759840;
 RA Kaneo T., Nakamura Y., Wolk C.P.,
 RA Matanabe A., Irliguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraiki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 RT cyanobacterium Anabaena sp. strain PCC 7120.";
 RL DNA Res. 8:205-213(2001).
 CC
 CC [3]
 CC SEQUENCE OF 24-324 FROM N.A.
 RX MEDLINE=96099685; PubMed=8525061;
 RA Zhang C.C., Hugenulin S., Frilly A.;
 RT "Analysis of genes encoding the cell division protein FtsZ and a
 RT glutathione synthetase homologue in the cyanobacterium Anabaena sp.
 RL PCC 7120.";
 CC Res. Microbiol. 146:445-455(1995).
 CC
 CC -1- CATALYTIC ACTIVITY: ATP + gamma-L-glutamyl-L-cysteine + glycine =
 CC ADP + phosphate + glutathione.
 CC
 CC -1- PATHWAY: Glutathione biosynthesis; second step.
 CC
 CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC GSH SYNTHASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U14408; AAA85527.1;
 DR EMBL: AP003594; BAB75558.1;
 DR EMBL: Z31371; CAA83242.1;
 DR HSSP: P04425; IGSH.
 DR InterPro: IPR004218; GSH-S_ATP.
 DR InterPro: IPR004215; GSH-S_Nterm.
 DR Pfam: PF02951; GSH-S_N.1.
 DR Pfam: PF02955; GSH-S_ATP; 1.
 KW Glutathione biosynthesis; Ligase; ATP-binding; Complete proteome.
 FT CONFLICT 50 50 W -> C (IN REF. 1).
 FT CONFLICT 243 243 M -> V (IN REF. 3).
 SQ SEQUENCE 324 AA: 35953 MW: 03C6E7E8481DD8D1 CRC64;

Query Match 51.6%; Score 33; DB 1; Length 324;
 Best Local Similarity 85.7%; Pred. No. 69;
 Matches 6: Conservative 0: Mismatches 1: Indels 0: Gaps 0:
 OY 6 AGORNFN 12
 DB 183 AGDRNFN 189

RESULT 50
 ID AMPC_PSEAE STANDARD; PRT; 397 AA.
 AC P24735;
 DT 01-MAR-1992 (Rel. 21, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Beta-lactamase precursor (EC 3.5.2.6) (Cephalosporinase).
 GN AMPC OR PA4110.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OC NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=91097488; PubMed=2125210;
 RA Lodge J.M., Minchin S.D., Piddock L.J.V., Busby S.J.W.;
 RT "Cloning, sequencing and analysis of the structural gene and
 RT regulatory region of the Pseudomonas aeruginosa chromosomal ampc
 RT beta-lactamase.";
 RL Biochem. J. 272:627-631(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Hickey C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Stoyer C.K., Brinkman F.S.L., Hutmagle W.O., Kowalik D.J., Lagrou M.,
 RA Gardner R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 RN [3]
 RP SEQUENCE OF 27-41.
 RX MEDLINE=94282208; PubMed=8012497;
 RA Michea-Hamzehpour M., Sanchez J.-C., Epp S.F., Paquet N., Hughes G.J.,
 RA Hochstrasser D.F., Pecheux J.-C.;
 RT "Two-dimensional polyacrylamide gel electrophoresis isolation and
 RT microsequencing of Pseudomonas aeruginosa proteins.";
 RL Enzyme Protein 47:1-8(1993).
 RN [4]
 RP SEQUENCE OF 1-3 FROM N.A.
 RC STRAIN=ATCC 15692 / PA01;
 RX MEDLINE=94010221; PubMed=8405939;
 RA Lodge J.M., Busby S.J.W., Piddock L.J.V.;
 RT "Investigation of the Pseudomonas aeruginosa ampr gene and its role
 RT at the chromosomal ampc beta-lactamase promoter.";
 RL FEMS Microbiol. Lett. 111:315-320(1993).
 CC -1- CATALYTIC ACTIVITY: A beta-lactam + H(2)O = a substituted beta-
 CC amino acid.
 CC -1- SUBCELLULAR LOCATION: Periplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE CLASS-C BETA-LACTAMASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X54719; CA38522.1; -
 DR EMBL: AE004827; AAG07497.1; -
 DR EMBL: X67095; CAA7469.1; -
 DR PIR: S13408; S13408.
 DR HSSP: P05364; 2BLT.
 DR MEROPS: S12.00W; -
 DR InterPro: IPR001466; Beta_lactamase.
 DR InterPro: IPR001586; Beta_lactamase_C.
 DR Pfam: PF00144; beta-lactamase; 1.

DR PROSITE: PS00336; BETA_LACTAMASE_C; 1.
 KW Hydrolyase; Antibiotic resistance; Periplasmic; Signal;
 KM Complete proteome.
 FT SIGNAL 1 26
 FT CHAIN 27 397
 FT ACT_SITE 90 90 BETA-LACTAMASE.
 FT BINDING 342 344 BY SIMILARITY.
 FT CONFLICT 397 397 SUBSTRATE (BY SIMILARITY).
 SQ SEQUENCE 397 AA; 43401 MW; C6341A53594BC261 CRC64;
 R -> A (IN REF. 1).
 Query Match 51.6%; Score 33; DB 1; Length 397;
 Best Local Similarity 50.0%; Pred. NO. 85;
 Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Search completed: May 16, 2003, 10:40:18
 Job time : 21 secs